



Characterisation of Anomalous Behaviour for Security in Deep-Edge Wireless Systems

Jiali Xu

► To cite this version:

Jiali Xu. Characterisation of Anomalous Behaviour for Security in Deep-Edge Wireless Systems. Computer Science [cs]. Lille University; Inria Lille, 2025. English. ⟨NNT : ⟩. ⟨tel-05420774⟩

HAL Id: tel-05420774

<https://hal.science/tel-05420774v1>

Submitted on 17 Dec 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC0 1.0 - Universal - International License



UNIVERSITÉ DE LILLE
INRIA

Doctoral School ED MADIS 631
University Department Centre Inria de l'Université de Lille

Thesis defended by **Jiali Xu**

Defended on **December 3, 2025**

In order to become Doctor from Université de Lille

Academic Field **Computer Science**

Speciality **IT and Applications**

Characterisation of Anomalous Behaviour for Security in Deep-Edge Wireless Systems

**Applied Machine Learning From On-Device Jamming Detection to
Collaborative Intelligence**

Thesis supervised by Valéria LOSCRÌ Supervisor
Romain ROUVOY Co-Supervisor

Committee members

<i>Referees</i>	Marcelo DIAS DE AMORIM	Senior Researcher at Sorbonne Université	
	Mališa VUČINIĆ	Senior Researcher at Inria Paris	
<i>Examiner</i>	Antoine GALLAIS	Professor at INSA Hauts-de-France	Committee President
<i>Guest</i>	Aline CARNEIRO VIANA	Senior Researcher at Inria Saclay	
<i>Supervisors</i>	Valéria LOSCRÌ	Senior Researcher at Inria Lille	
	Romain ROUVOY	Professor at Université de Lille, CNRS, Inria Lille	

COLOPHON

Doctoral dissertation entitled “Characterisation of Anomalous Behaviour for Security in Deep-Edge Wireless Systems”, written by Jiali Xu, completed on December 17, 2025, typeset with the document preparation system L^AT_EX and the yathesis class dedicated to theses prepared in France.



UNIVERSITÉ DE LILLE
INRIA

École doctorale ED MADIS 631

Unité de recherche Centre Inria de l'Université de Lille

Thèse présentée par Jiali Xu

Soutenue le 3 décembre 2025

En vue de l'obtention du grade de docteur de l'Université de Lille

Discipline **Informatique**
Spécialité **Informatique et Applications**

Caractérisation des Comportements Anormaux pour la Sécurité dans les Systèmes Sans-Fil Deep-Edge

Apprentissage Automatique Appliqué : de la Détection des
Interférences à l'Intelligence Collaborative

Thèse dirigée par Valéria Loscri directrice
Romain Rouvoy co-directeur

Composition du jury

<i>Rapporteurs</i>	Marcelo DIAS DE AMORIM	directeur de recherche au Sorbonne Université	
	Mališa VUČINIĆ	directeur de recherche à l'Inria Paris	
<i>Examineur</i>	Antoine GALLAIS	professeur à l'INSA Hauts-de-France	président du jury
<i>Invitée</i>	Aline CARNEIRO VIANA	directrice de recherche à l'Inria Saclay	
<i>Directeurs de thèse</i>	Valéria LOSCRI	directrice de recherche à l'Inria Lille	
	Romain ROUVOY	professeur à l'Université de Lille, CNRS, Inria Lille	

The Université de Lille neither endorses nor censures authors' opinions expressed in the theses: these opinions must be considered to be those of their authors.

Keywords: 5g security, machine learning, jamming attacks, anomaly detection, data augmentation, collaborative intelligence

Mots clés : sécurité 5g, apprentissage automatique, attaques de brouillage, détection d'anomalies, augmentation de données, intelligence collaborative

This thesis has been prepared at

Centre Inria de l'Université de Lille

Parc scientifique de la Haute-Borne
40, avenue Halley - Bât A - Park Plaza
59650 Villeneuve d'Ascq - France

☎ (+33)(0)3 59 57 78 00
✉ contact-lille@inria.fr
Web Site <https://www.inria.fr/>

The Inria logo is a stylized, red, cursive script of the word "Inria". The letters are fluid and connected, with a slight upward curve at the end of the word.

CHARACTERISATION OF ANOMALOUS BEHAVIOUR FOR SECURITY IN DEEP-EDGE WIRELESS SYSTEMS
Applied Machine Learning From On-Device Jamming Detection to Collaborative Intelligence**Abstract**

The proliferation of Fifth-generation (5G) networks is critical for modern society, yet their reliance on the open radio frequency spectrum makes them persistently vulnerable to jamming attacks. Existing detection methods often lack the scalability, cost-effectiveness, and robustness required for widespread deployment, and their performance is further hindered by the scarcity of anomalous data needed to train reliable machine learning models. This thesis addresses these challenges by proposing and validating a comprehensive, multi-stage framework for characterising and detecting anomalous behaviour, progressing from robust on-device sensing to scalable, collaborative intelligence. First, this work establishes the feasibility of a hardware-agnostic approach, demonstrating that robust jamming detection can be achieved using only native system logs from commercial User Equipment (UE) and base stations. A holistic data processing and evaluation pipeline (SHIELD) is introduced, proving its effectiveness against sophisticated, power-modulated jammers where conventional methods fail. Second, to overcome the critical limitations of data scarcity and poor model generalization, a tailored conditional Generative Adversarial Network (GANSec) is developed. A novel synthetic-only training strategy is shown to significantly enhance detector robustness, enabling models to maintain high performance under unseen network conditions. Finally, the solution is scaled to a system-wide level through a lightweight collaborative intelligence framework. This architecture leverages minimalist binary reporting from a multitude of UEs, which are aggregated to achieve network-wide situational awareness, including jammed area mapping and threat localization, with minimal communication overhead. The results of this research demonstrate a practical and effective end-to-end solution that advances the state of the art in wireless security. Serving as a foundational component of the EU-funded MLSysOps project, this work provides the core methodologies for a wireless network anomaly detection module aimed at ensuring resilience and trust in heterogeneous networks. By uniting on-device data collection with advanced machine learning for both data augmentation and collaborative sensing, this thesis presents a scalable and resilient paradigm for defending next-generation networks against physical layer threats.

Keywords: 5g security, machine learning, jamming attacks, anomaly detection, data augmentation, collaborative intelligence

CARACTÉRISATION DES COMPORTEMENTS ANORMAUX POUR LA SÉCURITÉ DANS LES SYSTÈMES SANS-FIL DEEP-EDGE**Apprentissage Automatique Appliqué : de la Détection des Interférences à l'Intelligence Collaborative****Résumé**

La prolifération des réseaux de cinquième génération (5G) est essentielle pour la société moderne, mais leur dépendance au spectre radiofréquence ouvert les rend vulnérables aux attaques par brouillage. Les méthodes de détection existantes manquent souvent de scalabilité, de rentabilité et de robustesse pour un déploiement à grande échelle, et leurs performances sont limitées par la rareté des données anormales nécessaires à l'entraînement de modèles d'apprentissage automatique fiables. Cette thèse aborde ces défis en proposant et en validant un framework complet et multi-étapes pour la caractérisation et la détection des comportements anormaux, progressant d'une détection robuste sur appareil à une intelligence collaborative et scalable. Premièrement, ce travail établit la faisabilité d'une approche agnostique du matériel, démontrant qu'une détection de brouillage robuste peut être réalisée en utilisant uniquement les journaux systèmes natifs des équipements utilisateurs (UE) et des stations de base commerciales. Un pipeline holistique de traitement et d'évaluation des données (SHIELD) est introduit, prouvant son efficacité contre des brouilleurs sophistiqués à modulation de puissance, là où les méthodes conventionnelles échouent. Deuxièmement, pour surmonter les limitations critiques de la rareté des données et de la faible généralisation des modèles, un réseau antagoniste génératif conditionnel (GANSec) est développé. Une nouvelle stratégie d'entraînement basée uniquement sur des données synthétiques améliore significativement la robustesse des détecteurs, permettant aux modèles de maintenir de hautes performances dans des conditions réseau inédites. Enfin, la solution est étendue à l'échelle du système grâce à un framework d'intelligence collaborative léger. Cette architecture s'appuie sur des rapports binaires minimalistes provenant d'une multitude d'UE, qui sont agrégés pour obtenir une connaissance situationnelle à l'échelle du réseau, incluant la cartographie de la zone brouillée et la localisation de la menace, avec une surcharge de communication minimale. Les résultats de cette recherche démontrent une solution de bout en bout, pratique et efficace, qui fait progresser l'état de l'art en matière de sécurité sans fil. Servant de composant fondamental au projet MLSysOps financé par l'UE, ce travail fournit les méthodologies de base pour un module de détection d'anomalies de réseau sans fil visant à garantir la résilience et la confiance dans les réseaux hétérogènes. En unifiant la collecte de données sur appareil avec des techniques d'apprentissage automatique avancées pour l'augmentation de données et la détection collaborative, cette thèse présente un paradigme scalable et résilient pour la défense des réseaux de nouvelle génération contre les menaces de la couche physique.

Mots clés : sécurité 5g, apprentissage automatique, attaques de brouillage, détection d'anomalies, augmentation de données, intelligence collaborative

Acknowledgement

The journey of completing this doctoral thesis has been a challenging yet immensely rewarding endeavor, and it would not have been possible without the guidance, support, and encouragement of many individuals to whom I owe my deepest gratitude.

First and foremost, I would like to express my sincere appreciation to my supervisor(s), Dr. Valéria Loscrì and Dr. Romain Rouvoy, for their invaluable mentorship, unwavering support, and insightful guidance throughout this research. Their expertise, patience, and constant encouragement were instrumental in shaping the direction of this work and my development as a researcher. I am profoundly grateful for the countless hours they dedicated to discussing ideas, reviewing manuscripts, and providing the intellectual freedom that allowed this thesis to come to fruition.

I am also deeply indebted to my co-authors and collaborators, whose contributions were vital to the publications that form the core of this thesis: Dr. Aya Moheddine, Dr. Alessandro Brighente, Dr. Mauro Conti, and Shuo Wang. Their collaborative spirit, insightful discussions, and shared dedication made the research process both productive and enjoyable.

My gratitude extends to all my colleagues at the Inria Centre at the University of Lille and the University of Padova. The stimulating and friendly research environment they fostered was a constant source of motivation and inspiration. I would also like to thank the members of my thesis committee for their time, valuable feedback, and constructive criticism, which have significantly improved the quality of this manuscript.

This research was made possible with support from the Horizon Europe research and innovation program of the European Union under grant agreement number 101092912 (project MLSysOps) and COST Action CA 22104 BEiNG-WISE. I am grateful for this financial support, which provided the resources necessary to conduct the experiments and disseminate the findings.

Finally, on a personal note, I wish to extend my heartfelt thanks to my family and friends. Their unconditional love, endless encouragement, and profound patience have been my bedrock of support throughout this long journey. To my partner, Wenhui, thank you for believing in me, especially during the most challenging times. This achievement is as much yours as it is mine.

Acronyms

Numbers | **A** | **B** | **C** | **D** | **E** | **F** | **G** | **I** | **K** | **L** | **M** | **N** | **O** | **P** | **Q** | **R** | **S** | **T** | **U** | **V** | **W**

Numbers

5GC 5G Core. 41, *Glossary*: 5GC

A

ADB Android Debug Bridge. 41, 45, 107, *Glossary*: ADB

AoA Angle of Arrival. 89

AoI Age of Information. 98, 99

AUC Area Under the Curve. 25

B

BER Bit Error Rate. 49

BLE Bluetooth Low Energy. 13

BPA Basic Probability Assignment. 93, 94

BPR Bad Packet Ratio. 56, 57

C

CDF Cumulative Distribution Function. 82

cGAN Conditional GAN. 71, 73

CN Core Network. 2, 12, 13, 41

CNN Convolutional Neural Networks. 27, 74, 75, 77, 82–85

CO Classification Oriented. 75, 76, 79, 81, 84, 85, 108

COTS Commercial Off-The-Shelf. 6, 7, 10, 35, 38, 41, 103, 104, 109

CQI Channel Quality Indication. 46, 49

CSI Channel State Information. 62, *Glossary*: CSI

D

DBSCAN Density-Based Spatial Clustering of Applications with Noise. 98, 100, 111

DDoS Distributed Denial-of-Service. 14

DoS Denial-of-Service. 2, 3, 15, 17, 31, 33, 41, 98

DST Dempster-Shafer Theory. 10, 93, 94, 98, 101

E

EC Embedded Conditional. 75, 76, 79, 81, 84, 85

eMBB enhanced Mobile Broadband. 15, *Glossary: eMBB*

EPC Evolved Packet Core. 2, 12, 15, *Glossary: EPC*

F

FDI False Data Injection. 97, 98, 100, 110, 111

FID Fréchet Inception Distance. 81, 82

G

GAN Generative Adversarial Network. 6, 7, 10, 32–34, 68, 70–76, 79, 80, 82, 84, 85, 105, 108, 109

GBDT Gradient Boosting Decision Tree. 57–60

GDPR General Data Protection Regulation. 5, 63

gNB Next-Generation Node B. xxv, 6, 10, 15, 17, 38, 39, 41, 44–48, 54, 55, 58, 59, 109, 111, 132, *Glossary: gNB*

I

IAT Inter-Arrival Time. 23

IDS Intrusion Detection Systems. 32

IoT Internet of Things. 1–5, 9, 11, 13–16, 22, 24, 26, 28–31, 33–37, 89, 90, 97, 98, 100, 101, 106, 110, 112

K

KDE Kernel Density Estimation. 54

KL DIVERGENCE Kullback-Leibler divergence. 72

KNN K-Nearest Neighbors. 57–59

KPI Key Performance Indicator. 22, 26, 27, 41, 46, 47

L

LSTM Long Short-Term Memory. 74, 77, 82, 84, 85, 107

M

MAB Multi-Armed Bandit. 21

MAC Medium Access Control. 45

MAE Mean Absolute Error. 50, 51

MAPE Mean Absolute Percentage Error. 50, 51

MEC Minimum Enclosing Circle. 94, 95, 98–101

MitM Man-in-the-Middle. 15

ML Machine Learning. 3–6, 9, 11, 26–35, 37, 39, 44, 47, 51, 56, 60–62, 64–66, 69, 98, 103–106, 108, 109, 111, 112

MLP Multi-Layer Perceptron. 74, 77, 80, 82, 85

MMD Maximum Mean Discrepancy. 81, 82

N

NAS Non-Access-Stratum. 15

NFV Network Functions Virtualization. 2, 12, 92, *Glossary: NFV*

NIDS Network Intrusion Detection Systems. 31
NR New Radio. 12
NSA Non-Standalone. 2, 12, 13, 15, *Glossary: NSA*
NSSAI Network Slice Selection Assistance Information. 15, *Glossary: NSSAI*

O

OFDM Orthogonal Frequency-Division Multiplexing. 19
O-RAN Open Radio Access Network. 2, 12, *Glossary: O-RAN*

P

PDoA Power Difference of Arrival. 89
PDR Packet Delivery Ratio. 23, 26, 28, 30, 49, 56, 57, 104
PII Personally Identifiable Information. 63–65, 69
PUCCH Physical Uplink Control Channel. 46, 49, 55
PUSCH Physical Uplink Shared Channel. 46, 49, 55

Q

QoS Quality of Service. 3, 15, 16, 20

R

RAN Radio Access Network. 12, 41, 46, 110, *Glossary: RAN*
RF Radio Frequency. 1, 2, 4, 9, 11, 15–18, 23, 32, 33, 37, 88
RFF Radio Frequency Fingerprinting. 36–38, *Glossary: RFF*
RIL Radio Interface Layer. 46
RLE Run-Length Encoding. 97, 98
RMSE Root Mean Squared Error. 51
RNF Random Forest. 56–60, 104
RNN Recurrent Neural Networks. 51, 74
ROC Receiver-Operating Characteristic. 25
RSRP Reference Signal Received Power. 46, 47, 49, 50, 52, 54, 77, 91
RSRQ Reference Signal Receiver Quality. 46, 49, 91
RSS Received Signal Strength. 20, 89, 90
RSSI Received Signal Strength Indication. 22, 28, 42, 43, 46, 49, 56, 57, 94, 104
RTSA Real-Time Spectrum Analyzer. 24

S

SA Standalone. 2, 13, *Glossary: SA*
SDN Software-Defined Networking. 2, 12, *Glossary: SDN*
SDR Software-Defined Radio. 4, 18, 37, 41, 131, *Glossary: SDR*
SFD Start-of-Frame Delimiter. 18, 19
SIN Space Information Networks. 88
SINR Signal-to-Interference-plus-Noise Ratio. 46, 49, 54, 55, 91
SMOTE Synthetic Minority Over-sampling Technique. 32
SNR Signal-to-Noise Ratio. 17, 22, 28, 42, 56, 57, 67
SVM Support Vector Machine. 27, 56–59, 104

T

TDoA Time Difference of Arrival. 88

ToA Time of Arrival. 94

U

UDP User Datagram Protocol. 132

UE User Equipment. 6, 10, 12, 15, 16, 38, 39, 41, 44–48, 54, 58, 59, 87, 89–101, 106, 110–112, 132, 133

URLLC Ultra-Reliable Low-Latency Communications. 15, 37, *Glossary*: URLLC

V

VAE Variational Autoencoder. 68, 70–73

W

WSN Wireless Sensor Networks. 30, 40, 63, 89

Contents

Abstract	xi
Acknowledgement	xiii
Acronyms	xv
Contents	xix
List of Tables	xxiii
List of Figures	xxv
Introduction	1
General Context: The 5G Era and its Security Implications	1
The Transformative Role of 5G Networks	1
A New Paradigm, A New Threat Landscape	2
The Enduring Threat of Jamming	2
The MLSysOps EU Project	3
Problem Statement and Research Gaps	3
The Scalability and Cost Barrier	4
The Robustness Gap	4
The Data Scarcity Bottleneck	5
Thesis Statement and Research Questions	5
Overview of the Proposed Methodology	6
Summary of Contributions	7
Publications and Research Outputs	8
Thesis Outline	9
1 Background and Related Work	11
1.1 The 5G and IoT Security Landscape	11
1.1.1 The 5G Architectural Paradigm and Its Inherent Vulnerabilities	12
1.1.2 The IoT Ecosystem and Resource-Constrained Security	13
1.1.3 Convergence and the Expanded Attack Surface	14
1.1.4 Protocol Layer Threats: A Case Study on Network Slicing Integrity	15
1.2 Radio Frequency Jamming as a Physical Layer Threat	16
1.2.1 Fundamentals of Jamming Attacks	17
1.2.2 A Taxonomy of Jamming Strategies	18
1.3 Paradigms for Anomalous Behaviour Detection	21

1.3.1	Signature-Based vs. Anomaly-Based Detection for Novel Threats	21
1.3.2	Limitations of Traditional and Hardware-Centric Detection Methods	22
1.3.3	Performance Evaluation of Detection Systems	24
1.4	Machine Learning Approaches for Jamming Detection	26
1.4.1	The Rationale for Machine Learning in Wireless Security	26
1.4.2	A Critical Review of Current ML-Based Detection Systems	27
1.4.3	Identified Research Gaps in Existing Literature	27
1.5	Towards Next-Generation Detection: On-Device and Collaborative Intelligence	29
1.5.1	The Imperative for On-Device Detection at the Network Edge	29
1.5.2	The Power of Collaborative Sensing in Distributed Networks	30
1.5.3	Addressing Data Scarcity: A Review of Public Datasets and Synthetic Data Generation	31
1.6	Summary and Thesis Motivation	33
2	On-Device Jamming Detection using Native System Logs	35
2.1	Motivation: A Scalable, Hardware-Agnostic Approach	35
2.1.1	Foundational Principles of Hardware-Based RF Fingerprinting	36
2.1.2	The Scalability and Economic Impasse in Massive Deployments	36
2.2	The SHIELD Framework	39
2.3	Experimental Testbed and Threat Model	40
2.3.1	Testbed Components	41
2.3.2	Threat Model	41
2.4	A Holistic Data Collection and Preprocessing Pipeline	44
2.4.1	Synchronized Data Acquisition from UE and gNB Logs	44
2.4.2	Feature Engineering from Native Logs	47
2.4.3	Handling Asynchronous and Missing Data via Interpolation	48
2.4.4	Temporal Feature Extraction via Sliding-Window Aggregation	51
2.5	A Rigorous Methodology for Evaluating Model Robustness	53
2.5.1	Analysis of UE Feature Distributions	54
2.5.2	Analysis of gNB Feature Distributions	55
2.6	Performance Evaluation	56
2.6.1	Benchmarking Existing Methods	56
2.6.2	Performance of the SHIELD Pipeline	57
2.7	Chapter Summary	59
3	Enhancing Detector Robustness with Generative Data Augmentation	61
3.1	The Challenge of Data Scarcity and Model Generalization	61
3.1.1	The Data Acquisition Bottleneck	62
3.1.2	Distribution Shift and the Failure of Generalization	65
3.1.3	The Inadequacy of Traditional Augmentation Techniques	66
3.1.4	The Case for Generative Models: A New Paradigm for Data Synthesis	69
3.2	The GANSec Framework for Wireless Time-Series Augmentation	73
3.2.1	Conditional GAN Fundamentals	73
3.2.2	Generator and Discriminator Design	74
3.2.3	Fidelity-focused vs. Task-oriented Generation	75
3.2.4	Tailoring GANSec for Wireless Data	76
3.3	Evaluation of the Synthetic-Only Training Strategy	78
3.3.1	Experimental Design and Baselines	78
3.3.2	Training Settings for Diversity-Fidelity Balance	80

3.3.3 Quantitative Analysis of Synthetic Data Quality	81
3.3.4 Cross-Scenario Evaluation of Anomaly Detector Robustness	83
3.4 Chapter Summary	85
4 A Framework for Collaborative Intelligence	87
4.1 Fundamental Limitations of On-Device Threat Detection	87
4.1.1 Inability to Assess Geographic Scope and Scale	88
4.1.2 The Inherent Incapability of Single-Point Source Localization	88
4.1.3 Resolving Ambiguity: Differentiating Malicious Attacks from Benign Faults	89
4.2 A Lightweight Collaborative Intelligence Architecture	90
4.2.1 Core Architectural Principles	90
4.2.2 System Components and Workflow	91
4.3 Fusion and Localization from Distributed Binary Data	93
4.3.1 Aggregating Unreliable Inputs with Evidence Theory	93
4.3.2 Jammer Localization using Boundary-Based Techniques	94
4.3.3 Deriving Fine-Grained Network Insights	95
4.4 In-Depth Analysis of Operational Challenges and Mitigations	97
4.4.1 Communication Overhead	97
4.4.2 Data Reliability and FDI Resilience	97
4.4.3 Impact of UE Density and Distribution	99
4.4.4 System Dynamics and UE Mobility	99
4.4.5 Handling Multiple Jamming Sources	100
4.5 Chapter Summary	100
5 Discussion and Conclusion	103
5.1 Synthesis of Contributions	103
5.2 Interpretation of Key Findings	104
5.2.1 The Efficacy of On-Device, Log-Based Detection	104
5.2.2 The Power of the Synthetic-Only Training Strategy	105
5.2.3 The Viability of Lightweight Collaborative Intelligence	106
5.3 Implications of the Research	106
5.3.1 Practical Implications	106
5.3.2 Methodological Implications	108
5.3.3 Theoretical Implications	108
5.4 Limitations of the Presented Work	109
5.4.1 Scope of Experimental Validation	109
5.4.2 Assumptions in the Generative and Collaborative Models	109
5.4.3 Lack of Quantitative Evaluation for the Collaborative Framework	110
5.5 Future Research Directions	110
5.5.1 Expanding Experimental Validation	110
5.5.2 Advancing the ML Methodologies	111
5.5.3 Hardening the Collaborative Framework	111
5.5.4 Moving from Detection to Autonomous Mitigation	112
5.6 Concluding Remarks	112
Bibliography	113
Glossary	131
Contents	135

List of Tables

1.1	Comparative Analysis of Jamming Strategies	20
1.2	Comparison between Signature-Based and Anomaly-Based Intrusion Detection Methods	23
1.3	Common Evaluation Metrics for Classifiers with Formulas and Security-Oriented Use Cases	25
2.1	Comparative Cost Analysis of RF Sensing and Attack Hardware	38
2.2	Feature set extracted from UE and gNB logs in a 10-minute session.	49
2.3	Quantitative comparison of interpolation methods using forward filling as a reference.	51
2.4	Comparative analysis of state-of-the-art jamming detection methods.	56
2.5	Hyperparameter tuning search space and optimal configurations for each ML model.	58
3.1	Qualitative Analysis of Commercial Software-Defined Radio (SDR) Devices from an Economic Perspective.	64
3.2	A Comparative Analysis of Traditional and Generative Data Augmentation Techniques.	68
3.3	A Comparative Overview of Key Generative Model Architectures.	71
3.4	Peak Performance Metrics on Unseen Scenario B.	83
4.1	Summary of Operational Challenges and Mitigation Strategies.	98

List of Figures

1	The MLSysOps framework architecture.	4
1.1	5G network slicing overview.	16
1.2	Principle registration procedure of 5G NAS protocol, elements encapsulated in messages are marked in blue, malicious Next-Generation Node B (gNB)'s interference is marked in red.	17
2.1	Architecture of the SHIELD Framework.	40
2.2	Experimental setup for data collection.	42
2.3	Impact of Power-Modulated Jamming on KPIs from iperf3.	44
2.4	End-to-End Synchronized Data Collection Flow in a 5G Infrastructure.	45
2.5	Comparative analysis of interpolation techniques.	50
2.6	KDE analysis comparing UE feature distributions.	54
2.7	KDE analysis comparing gNB feature distributions.	55
2.8	Model performance metrics on the UE dataset.	59
2.9	Model performance metrics on the gNB dataset.	59
3.1	GAN-based Data Augmentation for Anomaly Detection in Wireless Security: Workflow Overview.	73
3.2	EC- and CO-training mechanisms within GANSec framework.	78
3.3	Training Dynamics of the EC-GAN and CO-GAN.	81
3.4	CDF of FID Scores for Synthetic Samples over 100 Generation Iterations.	82
3.5	CDF of MMD Scores for Synthetic Samples over 100 Generation Iterations.	83
3.6	Impact of the Augmentation Ratio on the Accuracy Achieved on the Unseen Dataset B.	85
4.1	Overview of the Proposed Collaborative Jamming Detection Workflow.	92
5.1	Sequence diagram of the real-time jamming detection pipeline prototype, implemented on Android logs.	107

Introduction

This chapter introduces the research presented in this dissertation. It begins by establishing the broad technological and societal context of fifth-generation (5G) wireless networks, highlighting their transformative potential, the new security challenges they introduce, and the European research initiative, MLSysOps, that frames this work. It then narrows its focus to the specific problem of Radio Frequency (RF) jamming as a persistent physical layer threat. Subsequently, it presents the central thesis statement and the research questions that guide this work. An overview of the proposed three-stage methodology is provided, followed by a summary of the key contributions and a list of publications derived from this research. The chapter concludes with a detailed outline of the dissertation, providing a roadmap for the reader.

General Context: The 5G Era and its Security Implications

This section sets the broad stage for the research, establishing the critical importance of 5G networks and progressively narrowing the focus from the general threat landscape to the specific problem of RF jamming.

The Transformative Role of 5G Networks

The 5G of cellular networks represents a paradigm shift in global communications, engineered not merely as an incremental enhancement to mobile broadband but as a foundational platform for a new wave of societal and technological transformation. Unlike previous generations, which were primarily designed for consumer voice and data services, 5G is envisioned as a critical infrastructure component, providing the high-throughput, ultra-low-latency, and massively scalable connectivity required to underpin the digital economy of the 21st century. It becomes the critical enabler for a suite of transformative technologies that promise to reshape industries and daily life.

These technologies include the massive Internet of Things (IoT), where billions of diverse and resource-constrained endpoints, from environmental sensors to industrial actuators, will be seamlessly connected [252]. It also includes the development of autonomous systems, such as self-driving vehicles and unmanned aerial vehicles, which rely on instantaneous communication

for safe and efficient operation. Furthermore, 5G is poised to revolutionize critical industrial applications through concepts like Industry 4.0 [134], enabling smart factories, remote surgery, and real-time control of public utilities. The ambition of 5G is to serve as a unified connectivity fabric for these mission-critical services, positioning it as an indispensable element of national and global infrastructure whose reliability and security are of paramount importance.

A New Paradigm, A New Threat Landscape

The architectural innovations that empower 5G's transformative capabilities also introduce a more complex security landscape. The evolution from previous cellular generations is marked by a profound shift towards a purely software-based network paradigm, heavily reliant on technologies such as Network Functions Virtualization (NFV) and Software-Defined Networking (SDN). This softwarization, coupled with a strategic disaggregation of network components, creates a system that is more flexible, agile, and open than its monolithic predecessors. However, this openness and complexity significantly expand the attack surface, introducing novel vulnerabilities that demand a re-evaluation of traditional security models [164].

A key example of this new paradigm is the global initiative towards Open Radio Access Network (O-RAN). By advocating open and interoperable interfaces, O-RAN seeks to overcome the constraints of proprietary, vertically integrated architectures, thereby enabling the development of a more diverse and competitive multi-vendor ecosystem. While economically beneficial, this disaggregation raises new security challenges, as the existing specifications do not consistently follow a "security by design" paradigm, thereby exposing multiple interfaces to potential vulnerabilities exploitable by a wide range of adversaries [254, 128].

Further complicating the security posture is the phased nature of 5G deployment. Most operators have initially deployed 5G in a Non-Standalone (NSA) configuration, where the 5G radio interface is used but the network relies on the existing 4G Evolved Packet Core (EPC) for control and management functions. The profound security and privacy enhancements promised by 5G are primarily implemented within the new 5G Standalone (SA) Core Network (CN). Consequently, networks operating in NSA mode remain susceptible to a wide range of legacy attacks that target vulnerabilities in the 4G core. This creates a dangerous period of transitional insecurity, where networks are marketed as 5G but inherit 4G weaknesses, underscoring the need for security solutions that are effective in these hybrid, multi-generation environments.

The Enduring Threat of Jamming

Among the architectural complexities of 5G softwarization and the cross-layer vulnerabilities of the IoT, there exists a common, foundational point of weakness: the physical layer. All wireless systems, regardless of their specific protocols or applications, rely on the open and shared nature of the RF spectrum for communication. This shared medium is inherently susceptible to interference, making physical layer Denial-of-Service (DoS) attacks, most notably RF jamming, a

universal and highly effective threat vector [20, 189].

An adversary does not need to exploit complex software vulnerabilities or break sophisticated encryption to cause disruption. Instead, they only need to overpower the legitimate signal in the physical domain. The impact of a successful jamming attack is immediate and severe, manifesting as high packet loss, increased latency, and a significant rise in energy consumption as devices fruitlessly attempt retransmissions. In the context of 5G's mission-critical applications, the consequences are magnified. The degradation of Quality of Service (QoS) or a complete DoS condition could have devastating effects on latency-sensitive services such as autonomous vehicle control, remote surgical procedures, or industrial automation, where even a brief disruption in connectivity can lead to catastrophic failure [207]. This makes the study of jamming attacks not just relevant, but critical to ensuring the availability and reliability of future 5G-enabled services.

The MLSysOps EU Project

The research presented in this dissertation was fully funded by the EU Horizon Europe project MLSysOps (Machine Learning for Autonomic System Operation in the Heterogeneous Edge-Cloud Continuum). The project addresses the immense challenges of managing the loose federation of heterogeneous and distributed resources at an extreme scale, a task for which traditional human-in-the-loop management is unrealistic.

The vision of MLSysOps is to enable autonomic, efficient, and adaptive end-to-end system management and application orchestration on the dynamic edge-cloud continuum by using Machine Learning (ML) models. As depicted in Figure 1, its key objectives include delivering an open AI-ready framework, developing an AI architecture with retrainable ML models, and supporting green, resource-efficient, and trustworthy system operation while satisfying application QoS requirements. Figure 1a illustrates the high-level functional components, positioning the Anomaly Detection module as a core capability alongside Trust Assessment & Management. Figure 1b provides a more detailed view of the telemetry data pipeline, showing how the Anomaly Detection module consumes this telemetry data to identify potential threats.

The contributions of this thesis align directly with the project's goal of ensuring trustworthy and resilient system operation. The development of the ML-based anomaly detection module constitutes a core component of the project's strategy for enhancing communication security and trust across the continuum. This work provides the foundational mechanisms for detecting physical layer threats, which is a critical element for the autonomic management of secure and reliable networks.

Problem Statement and Research Gaps

The evolution of jamming attacks toward stealthy and adaptive strategies, coupled with the unique architectural and scale challenges of 5G and IoT networks, has exposed critical gaps in

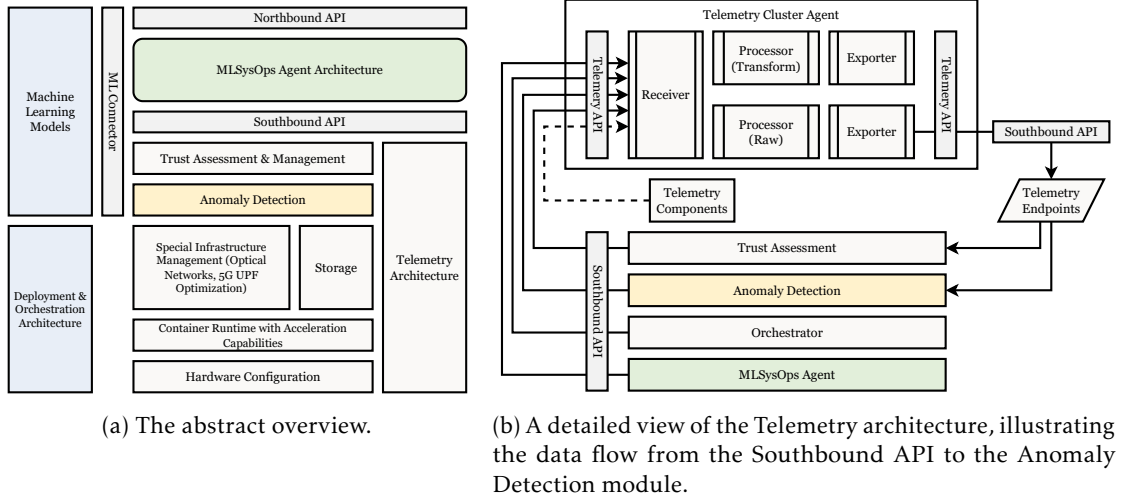


Figure 1: The MLSysOps framework architecture.

the state-of-the-art of wireless security. Existing detection methodologies are largely insufficient, creating a pressing need for new security solutions that are scalable, robust, and data-resilient. This thesis is motivated by three fundamental and interconnected research gaps that currently prevent the effective deployment of ML-based security in real-world wireless environments.

The Scalability and Cost Barrier

Traditional, hardware-centric methods for detecting physical layer attacks, such as those relying on dedicated spectrum analyzers, are fundamentally non-viable for securing the vast and heterogeneous IoT ecosystems [189, 142]. These solutions are characterized by prohibitive costs and a significant lack of scalability, rendering them economically and logistically infeasible for the ubiquitous deployment required. This creates a severe and unfavorable economic asymmetry between attackers and defenders. Effective jamming and spoofing attacks can be implemented using low-cost, commercially available hardware, such as Software-Defined Radio (SDR), which can be purchased for as little as \$150–\$500. In stark contrast, the high-precision RF analysis equipment required for reliable defense can cost tens or even hundreds of thousands of dollars per unit. This disparity creates a scalability-security-cost trilemma, where any attempt to scale a defense based on expensive, specialized hardware is an economically losing proposition. This gap establishes the clear and unavoidable necessity for a software-based, hardware-agnostic solution that can be deployed at massive scale without incurring crippling economic costs.

The Robustness Gap

The application of ML to wireless security has shown great promise, but the field suffers from a significant gap of real-world applicability. A vast majority of existing ML-based detection meth-

ods [149, 21, 25, 71, 88, 100] are developed and validated using data generated from network simulators like ns-3. While useful for initial testing, simulators inherently fail to capture the full complexity, unpredictability, and non-idealities of real-world wireless environments. Models trained exclusively on this sanitized, synthetic data are often brittle and poorly generalized. When deployed in a real network, they exhibit a significant drop in performance because they have not been trained to be robust to the noise and variability of the physical world.

This failure of generalization was empirically demonstrated in the research for this thesis. We tested benchmarked methods from the literature [238, 236, 71, 21, 100] which had reported accuracies of over 95% in their original studies. When tested against a sophisticated, real-world power-modulated jammer, their performance collapsed to as low as 36%. This robustness gap is driven by the statistical phenomenon of distribution shift between controlled training environments and dynamic operational environments. It highlights the critical need for models that are both developed and validated on real hardware against sophisticated threats. Furthermore, these models must be inherently resilient to the environmental changes they will inevitably encounter in deployment.

The Data Scarcity Bottleneck

The robustness gap is a symptom of a more fundamental problem: the data scarcity bottleneck. The development of any robust supervised ML model is predicated on the availability of a large-scale, high-quality, and accurately labeled training dataset. In the domain of wireless security, acquiring such a dataset is a formidable challenge, encumbered by a reinforcing system of physical, economic, and compliance constraints that form a bottleneck of data acquisition.

Physical barriers arise from the chaotic and non-stationary nature of the wireless medium itself, making it nearly intractable to collect a comprehensive and reproducible dataset [127, 185, 198]. Economic barriers include the high cost of specialized hardware for data collection and the substantial human capital required from security experts for accurate data labeling [48, 127, 185]. Finally, compliance barriers, such as global data privacy regulations like the General Data Protection Regulation (GDPR), impose stringent legal and ethical constraints on the collection and use of network traffic, which often contains sensitive personal information [240]. This fundamental scarcity of diverse, real-world anomaly data forces researchers to rely on unrealistic simulations and severely limits the ability of ML models to learn a generalizable representation of threats, preventing them from performing reliably when faced with unseen network conditions.

Thesis Statement and Research Questions

This dissertation argues that a practical, resilient, and cost-effective solution for detecting advanced physical layer jamming attacks in massive 5G and IoT networks can be achieved through a synergistic, three-stage framework. This framework combines lightweight on-device detection using native system logs, robustness enhancement via tailored generative data augmentation,

and scalable collaborative intelligence. We contend that this approach bridges the critical gap between academic research and real-world deployability.

To validate this thesis, this dissertation seeks to answer the following primary research questions:

RQ1: Is it feasible to perform robust on-device jamming detection against stealthy, sophisticated, power-modulated jamming threats using only native system logs on Commercial Off-The-Shelf (COTS) hardware, thereby overcoming the scalability and cost barriers of specialized equipment?

RQ2: Can tailored generative data augmentation, specifically using a conditional Generative Adversarial Network (GAN), effectively overcome the data scarcity bottleneck to improve the generalization and robustness of wireless anomaly detectors when faced with significant environmental distribution shifts?

RQ3: Can a scalable, network-wide situational awareness, including jammer localization, be achieved through a lightweight collaborative framework that fuses minimalist, binary data from a distributed population of devices, addressing the spatial ambiguity inherent in single-point detection?

Overview of the Proposed Methodology

The solution presented in this dissertation is structured in three stages. Each stage builds upon the last, progressively addressing the identified research gaps and constructing a complete, end-to-end framework for wireless security. This logical progression moves from proving a concept is possible, to making it robust, and finally to making it scalable and context-aware.

Stage 1: Foundational On-Device Detection (SHIELD)

The first stage of the research establishes the foundational feasibility of the thesis's core premise. It introduces the SHIELD (Scalable and Holistic Evaluation Framework for ML-Based 5G Jamming Detection) framework [258], an end-to-end pipeline for collecting native system logs from both the User Equipment (UE) and the base station (gNB). This real-world data is then used to train and, crucially, to rigorously evaluate the robustness of lightweight ML classifiers against a sophisticated, power-modulated jammer. By demonstrating high-accuracy detection using only pre-existing data streams on COTS hardware, this stage directly addresses the scalability and cost barrier, proving that effective physical layer security is possible without reliance on expensive, specialized equipment.

Stage 2: Enhancing Robustness via Generative Data Augmentation (GANSec)

Building upon the foundational on-device detector, the second stage confronts the critical problem of model generalization and the data scarcity bottleneck. It introduces GANSec [257], a tailored conditional GAN framework designed specifically to synthesize high-fidelity, diverse wireless time-series data. This stage pivots to a data-centric solution, proposing a synthetic-only training strategy. By training a detector exclusively on a rich dataset generated by GANSec, this work demonstrates a powerful method to overcome the lack of real-world anomaly data and create models that are inherently resilient to the distribution shifts commonly encountered in operational wireless environments.

Stage 3: Scaling to Network-Wide Collaborative Intelligence

The final stage of the methodology addresses the inherent spatial limitations of any isolated, single-point detector. It proposes a conceptual framework for collaborative intelligence [256] that scales the on-device solution to achieve network-wide situational awareness. This system is designed to aggregate minimalist, binary status reports from a vast, distributed population of devices. By fusing these simple inputs, the framework can construct a high-fidelity map of the geographic scope of an attack, resolve the ambiguity between malicious threats and benign faults, and estimate the physical location of the jammer, thereby transforming isolated detections into actionable, system-level intelligence.

Summary of Contributions

The main novel contributions of this thesis are as follows:

- A scalable, hardware-agnostic framework (SHIELD) for on-device jamming detection, empirically demonstrating that native system logs from COTS devices are a sufficient and rich source of telemetry for detecting sophisticated physical layer attacks with high accuracy (approximately 94%).
- A rigorous two-dataset evaluation methodology for assessing model robustness, which provides a more realistic measure of generalization performance against environmental distribution shifts than standard train-test splits.
- A tailored generative data augmentation framework (GANSec) for wireless time-series data, which addresses the critical data scarcity bottleneck in wireless security.
- Validation of a synthetic-only training strategy, demonstrating that a detector trained exclusively on GAN-generated data can achieve superior generalization (92.1% accuracy) on unseen environments compared to one trained on limited real-world data (78.1% accuracy).

- A conceptual framework for lightweight, scalable collaborative intelligence, which leverages minimalist binary reports from distributed devices to achieve network-wide situational awareness, including jammer localization, thereby overcoming the spatial ambiguity of single-point detection.

Publications and Research Outputs

The work conducted during this doctoral research has produced several scientific and technical outputs. In addition to peer-reviewed conference papers that form the foundation of the main chapters, the results have also been disseminated through public project deliverables and open-source software repositories.

Conference Proceedings

- J. Xu, V. Loscri. "Leveraging UE-Level Collaborative Intelligence for Scalable Jamming Detection in 5G Networks". IEEE DCOSS-IoT 2025, Jun 2025, Lucca, Italy.
- J. Xu, A. Moheddine, V. Loscri, A. Brighente, M. Conti. "SHIELD: Scalable and Holistic Evaluation Framework for ML-Based 5G Jamming Detection". 20th International Conference on Availability, Reliability and Security (ARES), Aug 2025, Ghent (BE), Belgium.
- J. Xu, S. Wang, V. Loscri, A. Brighente, M. Conti, R. Rouvoy. "GANSec: Enhancing Supervised Wireless Anomaly Detection Robustness through Tailored Conditional GAN Augmentation". ESORICS 2025 - 30th European Symposium on Research in Computer Security, Sep 2025, Toulouse, France.

Journal

- J. Xu, V. Loscri, R. Rouvoy. "Integrity Under Siege: A Rogue gNodeB's Manipulation of 5G Network Slice Allocation", submitted to Computers & Security, Elsevier, 2025.

Book Chapter

- A. Moheddine, J. Xu, V. Loscri, R. Rouvoy. "Chapter 2 - Evaluation and performance metrics for IoT security networks". Advanced Machine Learning for Cyber-Attack Detection in IoT Networks, Academic Press, 2025, Pages 27-50, ISBN 9780443290329.

Public Project Deliverables

- D3.3 - "Final Version of AI-ready MLSysOps Framework", Horizon Europe Project ML-SysOps, April 2025. (Contributor; Section 5, Mechanisms for AI-ready Networking Infrastructure.)
- D4.3 "Final Version of System Simulators", Horizon Europe Project MLSysOps, December 2024. (Contributor; Section 5, Network Simulation Environment.)

- D4.4 "Final Version of AI Architecture and ML Models", Horizon Europe Project MLSysOps, July 2025. (Contributor; Section 6, ML driven anomaly detection system to monitor 5G edge networks.)
- D5.2 "Final Integration and Evaluation Report", Horizon Europe Project MLSysOps, to be available in December 2025. (Contributor; Section 4, Wireless Network Management and Security at the Edge.)

Open-Source Code Repositories

- SHIELD, implements supervised learning models for detecting wireless jamming from native Android logs, GitHub repository. Available at: <https://github.com/jxu96/SHIELD>.
- Synchronized UE-gNB log & trace collector, Gitlab repository. Available at: <https://gitlab.inria.fr/fun-team/5g-trace-analyzer>.
- GANSec, GAN-based data augmentation training and evaluation framework, Github repository. Available at: <https://github.com/jxu96/GanSec>.
- DEMO on real-time jamming detection, based on Android logs, Github repository. Available at: https://github.com/jxu96/DEMO_Android_JD
- SHIELD on-device detection, extension on industrial device (Teltonika RUTX50), Github repository. Available at: https://github.com/jxu96/teltonika_usecase.
- Network simulation environment for MLSysOps project, GitHub repository. Available at: <https://github.com/jxu96/ExtNetSim>.

Thesis Outline

The remainder of this thesis is organized as follows:

Chapter 1: Background and Related Work This chapter establishes the theoretical and empirical foundations for the thesis. It first delineates the security landscape of 5G and IoT systems and provides a comprehensive review of RF jamming strategies, from elementary to advanced. The chapter then critically evaluates existing intrusion detection paradigms to establish the necessity of anomaly-based methods. Finally, it surveys the state-of-the-art in ML-based detection to identify the critical research gaps in data realism, feature engineering, and deployment models that motivate this work.

Chapter 2: On-Device Jamming Detection using Native System Logs This chapter presents the first contribution of the thesis: the SHIELD framework. It details the design of a scalable, hardware-agnostic system for on-device detection, including its realistic power-modulated jammer threat model, a holistic pipeline for collecting and processing native logs from both the UE and gNB, and a rigorous two-dataset methodology for evaluating model robustness. The chapter empirically establishes the foundational feasibility of performing accurate and resilient jamming detection directly on COTS devices.

Chapter 3: Enhancing Detector Robustness with Generative Data Augmentation This chapter addresses the critical challenge of model generalization by introducing GANSec, a novel framework for generative data augmentation. It pivots to a data-centric approach to solve the fundamental problems of data scarcity and distribution shift. The chapter details the design of a tailored conditional GAN for synthesizing high-fidelity wireless time-series data and validates a synthetic-only training strategy, demonstrating that this approach can create detectors with superior robustness to unseen environmental conditions.

Chapter 4: A Framework for Collaborative Intelligence This chapter addresses the spatial limitations of single-node detection by proposing a conceptual framework for scalable, network-wide intelligence. It details a lightweight architecture that leverages minimalist, binary reports from a distributed population of devices to achieve collective situational awareness. The chapter outlines mechanisms for robust data fusion using Dempster-Shafer Theory (DST) and for jammer localization using boundary-based geometric algorithms, providing a blueprint for transforming isolated data points into actionable, system-level intelligence.

Chapter 5: Discussion and Conclusion This final chapter synthesizes the contributions of the preceding chapters, reflecting on their practical, methodological, and theoretical implications. It provides a deeper interpretation of the key findings, such as the efficacy of log-based detection, the power of synthetic-only training, and the viability of lightweight collaboration. The chapter also presents a self-critical assessment of the work's limitations, including the scope of experimental validation and underlying model assumptions. Finally, it outlines a comprehensive roadmap for future research and provides concluding remarks on the long-term vision for resilient wireless security.

Background and Related Work

This chapter establishes the theoretical and empirical foundations of the thesis. It first delineates the architectural and security landscape of 5G and IoT systems to highlight vulnerabilities arising from their convergence. Building on this context, the discussion provides a comprehensive review of RF jamming as a critical physical layer threat, classifying attack strategies from elementary to advanced. The narrative then shifts to a critical evaluation of existing intrusion detection paradigms, establishing the necessity of anomaly-based methods for novel threats while detailing the limitations of traditional approaches. Subsequently, a survey of the state-of-the-art in ML-based jamming detection identifies significant research gaps in data realism, feature engineering, and deployment models. The chapter culminates by introducing on-device and collaborative intelligence as promising frontiers that address the challenge of data scarcity, thereby setting the stage for the novel contributions of this research.

1.1 The 5G and IoT Security Landscape

The contemporary wireless ecosystem is defined by the convergence of two transformative technologies: the fifth-generation (5G) of cellular networks and the IoT. While 5G furnishes the high-throughput, low-latency connectivity backbone, the IoT constitutes a massive proliferation of diverse, often resource-constrained, endpoints. This symbiotic relationship promises to unlock novel applications across industrial, commercial, and consumer domains, yet the architectural shifts inherent to 5G and the unique characteristics of the IoT ecosystem also introduce a complex, multi-layered security landscape [247, 154]. The result is a vastly expanded and deeply vulnerable attack surface, the understanding of which is essential to contextualizing the specific threats and defense mechanisms central to this thesis.

1.1.1 The 5G Architectural Paradigm and Its Inherent Vulnerabilities

The deployment of 5G networks represents a significant architectural evolution beyond previous cellular generations. Engineered not merely for enhanced mobile broadband, 5G serves as a foundational platform for mission-critical services requiring ultra-low latency and massive connectivity. This ambition has driven a fundamental redesign of both the Radio Access Network (RAN) and the CN, introducing powerful new capabilities alongside novel security challenges [167]. At its core, the transition to 5G is characterized by a profound shift towards softwarization, a disaggregation of network components, and a phased deployment that prolongs hybrid-generation vulnerability.

A 5G network architecture is broadly decomposed into three principal components. These are the UE (the end-user device), the RAN (which manages the wireless connection), and the CN (which handles data routing, session management, and service delivery) [5]. The 5G New Radio (NR) standard operates across two distinct frequency ranges: Frequency Range 1 (FR1), encompassing sub-6 GHz bands, and Frequency Range 2 (FR2), which utilizes millimeter-wave (mmWave) bands between 24 GHz and 100 GHz to achieve multi-gigabit-per-second data rates [1, 2]. This framework is being transformed by a move towards a purely software-based paradigm, heavily reliant on technologies such as SDN and NFV within the 5G core [5]. While this softwarization enables unprecedented flexibility and network upgrades via software, it also introduces significant system complexity. This complexity, coupled with the interconnection of diverse 3GPP and non-3GPP networks, creates new protection vulnerabilities and a substantially wider attack surface, necessitating a re-evaluation of security threats [246, 167, 3].

The global initiative towards O-RAN further complicates the 5G security posture. Historically, mobile networks relied on proprietary, vertically integrated solutions from a few vendors, a practice that fostered a market oligopoly characterized by "vendor lock-in". This dependency stifled innovation and created significant supply chain risks as operators were compelled to use a single vendor's equipment [167]. The O-RAN initiative seeks to dismantle this model by promoting open, interoperable interfaces between RAN components to foster a more diverse and competitive multi-vendor ecosystem. While commercially beneficial, this disaggregation and openness introduce new security liabilities. Indeed, a risk analysis of O-RAN specifications has already revealed medium to high security risks across numerous interfaces and components [126]. A principal cause is the inconsistent adherence to a "security by design" paradigm during development, which leaves potential vulnerabilities for a range of attackers, from external actors to malicious insiders, to exploit [167, 126, 128].

One of the most immediate and critical vulnerabilities in the current 5G landscape stems from its deployment modes. The transition from 4G to a full 5G infrastructure is a multi-year process, compelling most operators to initially deploy 5G in a NSA configuration. In NSA mode, the 5G NR provides the radio interface, but the network relies on the existing 4G EPC for control and management. This approach, while offering a faster path to market for 5G services, can create a false sense of security [179]. The profound security and privacy

enhancements promised by 5G are overwhelmingly implemented within the new 5G SA CN [246]. Consequently, networks operating in NSA mode remain susceptible to a wide range of legacy attacks targeting vulnerabilities in the 4G core. Research has confirmed that attacks such as the IMSI Leak [56], which exposes user identity and facilitates tracking, can be successfully transferred to and executed on commercial 5G NSA networks with no apparent mitigation [176, 246]. This creates a dangerous gap between the public perception of 5G's security and the reality of its transitional deployment. This period of transitional insecurity, where networks are marketed as 5G but inherit 4G weaknesses, represents a significant security debt. It underscores the need for solutions effective in these hybrid, multi-generation environments, rather than those assuming the idealized security posture of a pure SA architecture.

1.1.2 The IoT Ecosystem and Resource-Constrained Security

The IoT marks a paradigm shift in computing, extending network connectivity to a vast and heterogeneous array of physical objects, from simple environmental sensors to complex industrial actuators. The architectural framework of IoT systems is typically conceptualized in a layered model to manage data flow from the physical world to analytical applications. Although the exact number of layers can vary, a common and comprehensive framework consists of four: the Perception/Device Layer, the Network/Transport Layer, the Processing/Behavioral Layer, and the Application/Trust Layer [16, 165]. This layered structure, while logically sound, presents a multi-faceted security challenge, as vulnerabilities at any single layer can compromise the integrity of the entire system [52].

The most defining and pervasive challenge within the IoT ecosystem is the severe resource constraints of the end devices. The foundational **Perception Layer**, also the device or hardware layer, is where physical devices, sensors, and actuators interface with the environment to collect data and execute commands. The most severe resource constraints are present at this layer. IoT devices are often designed for low cost, small form factor, and long battery life, which imposes significant limitations on computational power, memory, and energy. A typical sensor node, operating with less than 100 kB of RAM and a low-frequency CPU, is still expected to perform security-critical functions like encryption and secure boot [232]. These constraints render the implementation of heavyweight cryptographic algorithms or sophisticated intrusion detection systems infeasible, leaving devices highly vulnerable to physical tampering and firmware compromise through memory corruption or control-flow hijacking.

Responsible for data transmission between the perception layer and higher-level systems, the **Network Layer** encompasses a wide array of communication protocols. These include short-range technologies like Bluetooth Low Energy (BLE) and ZigBee, and long-range options such as LoRa and Wi-Fi, often using messaging protocols like MQTT for cloud-edge communication. Security challenges at this layer range from protocol-specific vulnerabilities, such as MQTT hijacking [133, 195], to more general threats like traffic injection, spoofing, and replay attacks [24]. Supporting resource-constrained devices necessitates lightweight communication and security mechanisms

that can operate within the strict memory and power budgets of the IoT ecosystem [16].

The **Behavioral/Application Layer** monitors the system's high-level logic and consistency. It processes sensor-collected data and makes decisions, often involving cloud-based analytics and storage. Security at this layer focuses on detecting anomalies not apparent at lower levels, such as a sensor providing physically implausible data or an actuator behaving inconsistently with its intended function [8]. Modeling a device's expected behavior is crucial for detecting these subtle, context-aware attacks and preventing privilege abuse that might otherwise go unnoticed by lower-level packet filters [165, 264].

Finally, the **Trust/Policy Layer** provides an overarching architectural structure for reliable and secure operation across different domains. This layer is concerned with implementing concepts like zero-trust identity enforcement, continuous authentication, and federated policy control [14, 7, 35]. These mechanisms are vital in multi-tenant environments where devices may be owned and operated by different parties. However, implementing a true zero-trust architecture, which requires continuous auditing and fine-grained access control, is computationally intensive. Its application is often practically limited to the cloud edge, leaving embedded runtime environments at the far edge operating under a more vulnerable, implicit trust model [139, 241, 72, 263].

Securing this complex, multi-layered ecosystem demands a holistic, cross-layer perspective that simultaneously addresses vulnerabilities across firmware, networks, behavior, and trust controls. The inherent limitations of IoT devices are not merely an implementation detail but a fundamental driver of systemic insecurity.

1.1.3 Convergence and the Expanded Attack Surface

The convergence of 5G and the IoT is a present reality, with 5G positioned as the critical enabler for massive-scale IoT deployments. This symbiotic relationship, where 5G provides the ubiquitous, high-performance connectivity required by billions of IoT devices, creates an unprecedentedly large, heterogeneous, and interconnected technological ecosystem. While this convergence unlocks immense potential, it also fundamentally expands the attack surface, creating a security environment where the vulnerabilities of each domain amplify the risks in the other [101].

The security posture of this converged ecosystem is defined not by its strongest component, such as the highly secured 5G core, but by its weakest and most numerous elements: the resource-constrained IoT devices at the edge. While the 5G core is designed with advanced, software-defined security mechanisms [3, 4], the IoT landscape is populated by billions of devices that are computationally and energetically incapable of supporting robust security measures by design [16]. When these fundamentally insecure endpoints are connected en masse via the 5G backbone, they become a systemic liability. A single compromised sensor or actuator, while seemingly insignificant, can serve as a beachhead for larger-scale attacks [208]. It can be used to inject malicious data [106], participate in a Distributed Denial-of-Service (DDoS) attack [31], or,

as is particularly relevant to this thesis, act as a tool for physical layer disruption [243, 217]. This "weakest link amplification effect" elevates the importance of on-device security from a localized concern to a systemic necessity for the stability and integrity of the entire 5G ecosystem.

Amidst the architectural complexities of 5G softwarization and the cross-layer vulnerabilities of the IoT, a common, foundational point of weakness exists: the physical layer. Both 5G and IoT systems, regardless of their specific protocols or applications, rely on the open and shared nature of the wireless RF spectrum for communication. This shared medium is inherently susceptible to interference [189, 164, 207]. An adversary does not need to exploit complex software vulnerabilities or break sophisticated encryption to cause disruption. They only need to overpower the legitimate signal in the physical domain. This makes physical layer DoS attacks, most notably RF jamming, a universal and highly effective threat vector that can incapacitate wireless services irrespective of security measures implemented at higher layers of the protocol stack [237]. This universal vulnerability renders the study of jamming attacks not just relevant, but critical to ensuring the availability and reliability of future 5G-enabled IoT services.

1.1.4 Protocol Layer Threats: A Case Study on Network Slicing Integrity

Beyond the physical layer, the sophisticated architecture of 5G introduces new vulnerabilities at the protocol level. A prime example is the threat to the integrity of network slicing, a key 5G feature allowing the creation of multiple virtual networks on a single physical infrastructure (see Figure 1.1). Each slice is tailored with specific QoS parameters to serve different applications, such as enhanced Mobile Broadband (eMBB) or Ultra-Reliable Low-Latency Communications (URLLC).

This protocol-level vulnerability is particularly pressing in current 5G NSA deployments, which depend on the existing 4G/LTE EPC. This architectural dependency means that 5G NSA inherits security weaknesses from the previous generation [49, 102, 123]. Research has already demonstrated that attackers can exploit misconfigurations and weak integrity protection in commercial LTE networks to perform sophisticated attacks [103, 203, 187]. The threat of fake base stations, which act as malicious intermediaries, is a well-established problem in 4G/LTE that persists in the 5G era [202, 218, 205]. Because 5G NSA uses the 4G control plane for initial attachment and session management, it remains susceptible to these inherited Man-in-the-Middle (MitM) attacks. This susceptibility allows adversaries to intercept and manipulate pre-authentication traffic, compromising critical procedures like network slice allocation.

Recent research has demonstrated a critical vulnerability in the network slice allocation process that a rogue gNB can exploit by acting as a MitM attacker. This threat model assumes an attacker can deploy a fake base station and leverage misconfigurations, such as the optional encryption of Non-Access-Stratum (NAS) messages, to compromise the integrity of the UE registration procedure. By inferring or intercepting key security parameters, the rogue gNB can alter the Network Slice Selection Assistance Information (NSSAI) [6] requested by the UE during its initial network connection. Figure 1.2 illustrates the full landscape of such a NAS-targeted

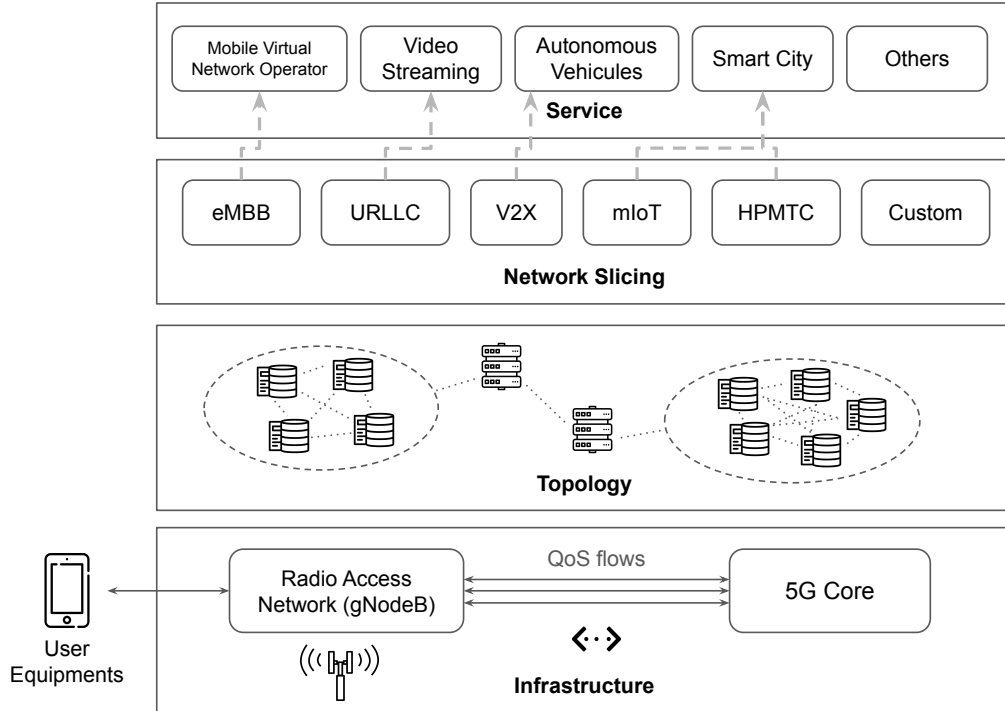


Figure 1.1: 5G network slicing overview.

attack.

The impact of such an attack is a subtle but significant form of service degradation. The attacker can force the UE to be allocated a non-optimal or geographically distant network slice without triggering any alarms in the 5G core. A proof-of-concept experiment has shown this can lead to increased latency and a considerable reduction in bandwidth, severely impacting the performance of latency-sensitive applications. This type of attack is particularly insidious because the resulting QoS degradation can easily be mistaken for legitimate network congestion or a system fault, making detection extremely challenging. This case study highlights the critical need for robust, end-to-end security measures that extend beyond physical layer defense to protect the integrity of core 5G functionalities.

1.2 Radio Frequency Jamming as a Physical Layer Threat

Having established the broad security landscape of 5G and IoT systems, this section now narrows its focus to the specific threat vector at the core of this thesis: RF jamming. As a physical layer attack, jamming represents a fundamental threat to the availability of any wireless communication system. Its efficacy lies in its simplicity and directness, targeting the medium of communication itself rather than complex protocols or software. This section defines the

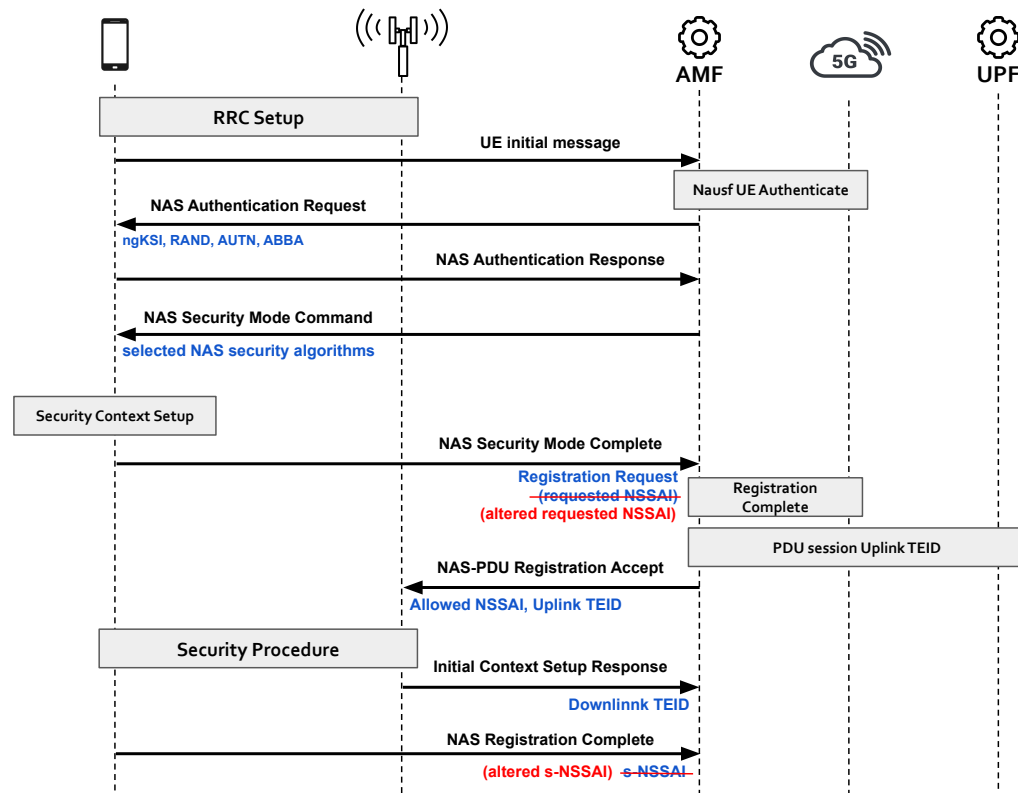


Figure 1.2: Principle registration procedure of 5G NAS protocol, elements encapsulated in messages are marked in blue, malicious gNB's interference is marked in red.

fundamentals of jamming, provides a systematic classification of common strategies, and traces their evolution toward more intelligent, adaptive, and stealthy forms that challenge traditional detection and mitigation paradigms.

1.2.1 Fundamentals of Jamming Attacks

At its core, jamming is the deliberate disruption of wireless communications through the transmission of interfering RF signals. A jammer's primary objective is to degrade the Signal-to-Noise Ratio (SNR) at a legitimate receiver to a point where the incoming signal can no longer be successfully decoded [20]. This malicious intent distinguishes jamming from unintentional interference, which may result from other nearby transmitters operating in the same frequency band. Jamming is an act specifically designed to cause a DoS condition at the physical layer [83, 181]. The impact of a successful jamming attack is immediate and severe, manifesting as high packet loss rates, increased communication latency, and a significant rise in energy consumption as legitimate devices fruitlessly attempt retransmissions [59]. In critical applications, this disruption can have devastating consequences for emergency services, public safety networks,

and industrial control systems [11].

A particularly concerning aspect of jamming is its accessibility. Unlike sophisticated cyberattacks that demand deep expertise, effective jamming can be executed using low-cost, commercially available hardware [20]. The proliferation of affordable SDRs has dramatically lowered the barrier to entry, enabling adversaries with minimal programming skills to create and deploy highly customized and effective jammers [128]. This combination of high impact and low execution cost makes RF jamming a pervasive, practical, and highly realistic threat to the reliability of modern wireless networks.

1.2.2 A Taxonomy of Jamming Strategies

Jamming attacks are not monolithic. They encompass a range of strategies that vary in sophistication, energy efficiency, and stealthiness. These strategies can be broadly classified into proactive and reactive jammers, with further subdivisions based on their specific behavior [83]. This taxonomy reveals an evolutionary trend from brute-force, easily detectable methods toward more intelligent and surreptitious techniques.

Proactive Jammers

Proactive Jammers are characterized by their indiscriminate transmission of interfering signals, operating regardless of whether legitimate communication is taking place [114]. This category includes several distinct types:

- **Constant Jammer:** In this most straightforward form of jamming, the attacker continuously transmits a high-power signal, which can consist of random bits or simple noise, to saturate the communication channel [28]. While highly effective at disrupting all communication within its range, this approach is extremely energy-intensive, and its persistent, unchanging signature makes it relatively easy to detect using simple energy-sensing techniques [181, 188].
- **Deceptive Jammer:** A slightly more sophisticated proactive strategy, the deceptive jammer continuously transmits signals mimicking legitimate network packets, such as preambles and Start-of-Frame Delimiter (SFD) [28]. The goal is to trick legitimate receivers into a "receive" state, locking them into processing the fake packet and rendering them unable to receive actual data [114]. This method is stealthier than a constant jammer because the interfering signal appears as valid traffic, but it remains highly inefficient in terms of energy consumption [38, 20].
- **Random Jammer:** This type of jammer attempts to balance effectiveness with energy conservation and stealth. It alternates between an active jamming state and a dormant state, with the durations of these periods being fixed or randomized [114]. During its active phase, it can behave as either a constant or a deceptive jammer. By periodically

ceasing transmission, the random jammer consumes less power and is harder to detect than a continuous jammer due to its intermittent presence on the channel [59, 164].

Reactive Jammers

Reactive Jammers represent a significant step up in intelligence and efficiency. Unlike proactive jammers, a reactive jammer first monitors the wireless channel for activity and only begins transmitting its interfering signal upon detecting the start of a legitimate transmission [114]. This targeted approach offers several advantages to the attacker. It is extremely energy-efficient, as power is expended only when an attack is guaranteed to be effective [59, 206]. Furthermore, it is far more difficult to detect. The jammer's signal is brief and always co-incident with a legitimate signal, making it hard to distinguish from normal collisions or poor channel conditions [58]. A common implementation is the SFD jammer, which listens for the SFD of a packet and then transmits a short burst of energy just in time to corrupt the payload, ensuring the packet is dropped at the receiver with minimal effort [259].

Table 1.1 provides a consolidated comparison of these fundamental jamming strategies, highlighting the trade-offs an attacker makes between effectiveness, energy cost, and detectability.

This classification clearly illustrates an evolutionary path in jamming technology. The progression from brute-force constant jammers to surgical reactive jammers is driven by an underlying strategic calculus: the adversary is engaged in an economic optimization problem. They seek to maximize the disruption caused per unit of energy expended while simultaneously minimizing the probability of being detected and neutralized. This economic view of jamming explains the clear trend towards more intelligent, efficient, and stealthy attack strategies. Consequently, any effective defense mechanism must disrupt this calculus by increasing the cost to the attacker, either by dramatically improving the probability of detection or by forcing the attacker to expend more energy for the same level of disruption.

Advanced Adaptive/Smart Jammers

The threat landscape is not static. Adversaries are continually developing more sophisticated jamming techniques that move beyond the basic taxonomy of proactive and reactive models. These advanced or smart jammers are characterized by their adaptability, precision, and their explicit design to circumvent countermeasures, representing a more realistic and formidable threat in modern wireless environments [268].

One key evolution is the development of targeted jammers. Rather than indiscriminately corrupting entire data packets, these advanced adversaries focus their limited energy budget on the most critical and vulnerable components of a wireless transmission. For example, in systems like Orthogonal Frequency-Division Multiplexing (OFDM) that rely on pilot symbols for accurate channel estimation, an attacker can specifically jam only these pilot symbols. This small, targeted attack can completely disrupt the channel estimation process, rendering the entire data transmission undecodable [211, 51]. Similarly, an attacker can choose to jam only

Jammer Type	Operating Principle	Energy Efficiency	Stealth / Detectability	Primary Impact
Constant	Continuously transmits a high-power noise or random signal.	Very Low	Low (Easy to detect via energy sensing).	Complete channel saturation; prevents all transmissions.
Deceptive	Continuously transmits legitimate-looking packets to lock receivers.	Very Low	Medium (Harder to distinguish from normal traffic).	Forces receivers into an unproductive listening state.
Random	Alternates between active jamming and sleeping states.	Medium	Medium (Intermittent nature complicates detection).	Intermittent disruption of communication; degrades QoS.
Reactive	Monitors the channel and jams only upon detecting a legitimate transmission.	High	High (Brief, targeted signal is difficult to distinguish from noise/collisions).	Causes targeted packet loss and forces retransmissions.
Advanced (smart)	Targets critical control signals (e.g., pilots, ACKs); may modulate power.	Very High	Very High (Surgical, low-power, and evades localization).	Disrupts core protocol functions; undermines countermeasures.

Table 1.1: Comparative Analysis of Jamming Strategies

the acknowledgement (ACK) packets. This causes the sender to believe the data packet was lost and initiate a costly retransmission, effectively halving the throughput with minimal energy expenditure [97, 189, 268]. By targeting specific weaknesses in the communication protocol, advanced jammers can exert a significant impact with minimal energy expenditure, while their short, surgical bursts render them particularly challenging to detect [73].

Another significant advancement is the emergence of power-modulated jammers. A primary countermeasure against jamming is to locate the physical source of the attack. Many localization techniques rely on measuring the Received Signal Strength (RSS) from multiple points to triangulate the jammer's position [231]. Power-modulated jammers are specifically designed to defeat this countermeasure. By dynamically and randomly varying their transmission power over time, they introduce high uncertainty into RSS measurements. This corruption of input data renders distance estimations unreliable and can make the localization process completely ineffective, even in environments with very little signal interference [159]. This strategy represents a fundamental shift in the adversarial mindset. The attack is no longer just on the availability of the communication channel but on the integrity of the defense mechanism itself. It is an

anti-forensic tactic, designed not only to deny service but to actively prevent attribution and response. This fundamentally alters the security challenge from a simple attack-detect-respond cycle to a more complex adversarial contest, where the detection and mitigation processes are themselves targets of the attack.

Finally, the most sophisticated adversaries employ adaptive jamming strategies. Real-world networks are not static. Legitimate nodes react to persistent interference by employing counter-measures such as switching frequencies, adapting data rates, or re-routing traffic [268]. A static jammer would quickly be rendered ineffective by these responses. A truly effective adversary must therefore also be adaptive, monitoring the network's response and dynamically adjusting its own strategy to maintain disruption [20]. This could involve changing the attack target, switching jamming protocols, or adjusting power levels in real-time. Researchers have explored the use of reinforcement learning by jammers to learn the most effective adaptive strategies in dynamic environments [118, 37, 64]. For instance, recent work by Bout et al. [37] has demonstrated a Multi-Armed Bandit (MAB)-based attack that can dynamically learn and follow a victim's channel hopping pattern in real-time, effectively countering this common defense mechanism. The rise of these intelligent, adaptive jammers makes it clear that static defense mechanisms are no longer sufficient. The defense must be equally dynamic and intelligent to counter an adversary that learns and adapts.

1.3 Paradigms for Anomalous Behaviour Detection

The evolution of jamming towards stealthy, adaptive, and intelligent strategies necessitates a corresponding evolution in detection methodologies. Simple, static defenses are insufficient to counter a dynamic threat. This section establishes the foundational paradigms for detecting such anomalous behavior in wireless networks. It begins by contrasting the two dominant approaches, signature based and anomaly based detection, to justify the focus on the latter for novel threats. It then critically examines the limitations of conventional methods for anomaly detection. Finally, it establishes a rigorous, quantitative framework for evaluating the performance of any proposed detection system, providing the formal metrics used throughout this thesis.

1.3.1 Signature-Based vs. Anomaly-Based Detection for Novel Threats

The field of intrusion detection is broadly divided into two fundamental paradigms: signature-based detection and anomaly-based detection. The choice between them represents a strategic trade-off between precision against known threats and the ability to identify novel ones.

- **Signature-Based Detection**, also known as misuse detection, operates like an antivirus program for network traffic. It maintains a database of known attack patterns, or "signatures," which can be specific byte sequences, malicious IP addresses, file hashes, or well defined patterns of malicious behavior [32]. When incoming network traffic is analyzed, it

is compared against this database. If a match is found, an alert is triggered. The primary strength of this approach lies in its high accuracy and extremely low rate of false positives for known threats. It is computationally efficient and provides clear, actionable alerts about the specific nature of the detected threat [132]. However, its core limitation is also its defining characteristic: it is fundamentally incapable of detecting new, unknown, or "zero day" attacks for which no signature exists. This inadequacy makes it an insufficient defense against the rapidly evolving jamming strategies discussed previously.

- **Anomaly-Based Detection**, in contrast, does not look for known bad behavior. Instead, it focuses on identifying deviations from established normal behavior. This process involves two phases. First, the system is trained on a dataset representing normal, legitimate network operation to build a statistical or machine learning model of this "normal" baseline. Second, the system continuously monitors live traffic, comparing it against the learned baseline. Any activity that deviates significantly from the established norm is flagged as a potential anomaly or intrusion [32]. The paramount advantage of this paradigm is its inherent ability to detect novel, zero day, and previously unseen attacks, as it does not rely on prior knowledge of specific attack patterns. This makes it the essential approach for countering sophisticated and adaptive threats. The principal challenge, however, is its tendency to produce a higher rate of false positives [107]. Benign but statistically rare events, such as a user logging in at an unusual time, can be misidentified as malicious, leading to alert fatigue for security analysts.

The distinct strengths and weaknesses of these two paradigms indicate that they are complementary. Many modern security architectures employ a hybrid approach, using signature-based detection as a fast and efficient first line of defense against common, known threats, while leveraging anomaly-based detection to identify novel and more sophisticated attacks that evade the signature database [197]. For the problem of detecting advanced adaptive jammers, which are by definition novel and designed to evade simple detection, the anomaly-based paradigm is not just an alternative, but a necessity.

1.3.2 Limitations of Traditional and Hardware-Centric Detection Methods

Within the paradigm of anomaly-based detection, the earliest and most straightforward approaches are traditional, non-machine learning methods that rely on monitoring network metrics and using specialized hardware. While conceptually simple, these methods suffer from fundamental limitations in adaptability, scalability, and cost that render them inadequate for the dynamic 5G and IoT environment.

Traditional metric-based detection techniques operate by monitoring a selection of Key Performance Indicators (KPIs) and flagging an anomaly when one or more of these metrics cross a predefined, static threshold [92]. Common metrics used for jamming detection include physical layer indicators like Received Signal Strength Indication (RSSI) and SNR, and MAC

Criterion	Signature-Based Detection	Anomaly-Based Detection
Detection Principle	Matches traffic against a database of known attack patterns (signatures).	Identifies deviations from a learned baseline of normal behavior.
Primary Strength	High accuracy and low false positive rate for known attacks.	Ability to detect novel, unknown, and zero-day attacks.
Primary Weakness	Completely ineffective against zero-day or polymorphic attacks.	Higher potential for false positives; benign anomalies can trigger alerts.
Effectiveness vs. Zero-Day	None. An attack must be known to have a signature.	High. Detection is based on behavior, not prior knowledge of the attack.
Typical False Positive Rate	Low.	Higher, requires careful tuning of the baseline and detection thresholds.
Maintenance Requirement	Requires constant and timely updates to the signature database.	Requires initial training to establish a robust baseline and periodic retraining.

Table 1.2: Comparison between Signature-Based and Anomaly-Based Intrusion Detection Methods

or application layer indicators like Packet Delivery Ratio (PDR) and packet Inter-Arrival Time (IAT) [74, 111, 19, 25]. The central flaw in this approach is its reliance on manually set, fixed thresholds [191]. Wireless environments are inherently dynamic. For example, signal strength fluctuates, and network load varies. Benign phenomena such as network congestion, poor link quality due to environmental factors, or device mobility can cause these metrics to behave in ways that are indistinguishable from a jamming attack when viewed through the lens of a simple threshold. This leads to a high rate of false alarms and poor overall detection accuracy. The process of manually tuning these thresholds is described in the literature as "arduous and error-prone," and it becomes exponentially more complex as more metrics are added in an attempt to improve accuracy [25]. This approach lacks the adaptability to handle the complexity and dynamism of modern wireless networks and sophisticated jammers.

Hardware-centric detection methods rely on specialized equipment, such as spectrum analyzers, to monitor the RF environment. A conventional swept-tuned spectrum analyzer measures power within a given bandwidth by sweeping across a frequency span [169]. While useful for analyzing static signals, this method has severe limitations when applied to the detection of modern jamming attacks in dynamic wireless environments like 5G and Wi-Fi. Because it sweeps across frequencies sequentially, it can easily miss signals that are intermittent, bursty, or that hop between frequencies. These characteristics are common to both legitimate modern

communication protocols and advanced jammers. An intermittent interferer that is hidden within a more powerful, legitimate downlink signal may contribute only minimally to the total power measured in a given sweep and thus go completely unnoticed. While more advanced Real-Time Spectrum Analyzers (RTSA) can capture and visualize these fast, time-varying signals more effectively, they represent a costly and specialized solution. Deploying such dedicated hardware across thousands or millions of low-cost IoT devices is economically and logistically infeasible [265]. Furthermore, these hardware-based solutions, like their traditional metric-based counterparts, often lack the flexibility to adapt to new and unknown attack strategies without requiring costly physical upgrades or firmware updates [269]. Their cost, lack of scalability, and limited adaptability make them unsuitable as a general-purpose solution for jamming detection in large-scale IoT networks. The problem of jamming detection has surpassed a complexity that can be solved by simple, human-defined rules or brute-force hardware analysis. The intricate, non-linear relationships between dozens of network parameters and the presence of a sophisticated jammer necessitate a move towards methods that can automatically learn these complex patterns from data.

1.3.3 Performance Evaluation of Detection Systems

To move beyond the limitations of traditional methods and develop more sophisticated detection systems, it is imperative to have a rigorous and standardized framework for evaluating their performance. In the context of this thesis, jamming detection is framed as a binary classification problem: for any given period of observation, the system must classify the network state as either "Normal" or "Jammed." The performance of such a classifier is quantitatively assessed using a set of well-defined metrics derived from a confusion matrix.

A confusion matrix is a table that summarizes the performance of a classification model by comparing its predictions against the ground truth. It provides a detailed breakdown of correct and incorrect classifications, which is essential for understanding a model's behavior, particularly in the context of imbalanced datasets common in security applications. The matrix is composed of four fundamental values:

- **True Positives (TP):** The number of instances correctly classified as belonging to the positive class (i.e., "Jammed" state correctly identified as "Jammed").
- **True Negatives (TN):** The number of instances correctly classified as belonging to the negative class (i.e., "Normal" state correctly identified as "Normal").
- **False Positives (FP):** The number of instances incorrectly classified as belonging to the positive class (i.e., "Normal" state incorrectly identified as "Jammed"). This is also known as a false alarm.
- **False Negatives (FN):** The number of instances incorrectly classified as belonging to the negative class (i.e., "Jammed" state incorrectly identified as "Normal"). This is also known

as a missed detection.

From these four values, several key performance metrics are calculated to provide a comprehensive assessment of a model's effectiveness. The choice of which metric is most important depends on the specific application and the relative costs of different types of errors. For instance, in security, a false negative (missing an attack) is often considered far more costly than a false positive (a false alarm).

The following Table 1.3 formally defines the core classification metrics that will be used to evaluate the models developed in this thesis. This framework establishes the methodological rigor required for a credible assessment of the proposed contributions and provides a standard against which they will be measured.

Metric	Mathematical Formula	Interpretation and Use Case
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	The overall proportion of correct classifications. It is a simple, intuitive metric but can be highly misleading on imbalanced datasets where a model can achieve high accuracy by simply predicting the majority class.
Precision	$\frac{TP}{TP + FP}$	Of all the instances the model predicted as positive, what proportion were actually positive? This metric is critical when the cost of a false positive is high (e.g., triggering a costly and disruptive response to a non-existent attack).
Recall	$\frac{TP}{TP + FN}$	Of all the actual positive instances, what proportion did the model correctly identify? Also known as Sensitivity or True Positive Rate (TPR), this is arguably the most critical metric in security, where the cost of a false negative is extremely high.
F1-Score	$\frac{2TP}{2TP + FP + FN}$	The harmonic mean of Precision and Recall providing a balanced measure of a model's performance, especially useful when there is an uneven class distribution and there is a need to balance false positives and false negatives.
ROC/AUC	N/A	The Receiver-Operating Characteristic (ROC) curve plots the trade-off between the true positive rate and false positive rate across classification thresholds. The Area Under the Curve (AUC) metric summarizes this curve, measuring overall class separation. Perfect discrimination scores 1.0, while random guessing scores 0.5.

Table 1.3: Common Evaluation Metrics for Classifiers with Formulas and Security-Oriented Use Cases

This evaluation highlights a crucial aspect of anomaly detection: the inherent tension between detecting all true threats and avoiding false alarms. A system that is highly sensitive (high recall) may flag many benign events as malicious (low precision), leading to a high false positive rate. The key to successful anomaly detection, therefore, is not merely observing a deviation from a baseline, but having sufficient context to correctly interpret the cause of that deviation. A single metric, like a drop in PDR, could be caused by either malicious jamming or benign network congestion. A system that lacks the broader context to distinguish between these two causes is doomed to a high rate of false positives. This points toward the need for detection systems that can fuse information from multiple sources. Such a multi-modal, context-aware view of the network is necessary to reduce the ambiguity that plagues simpler anomaly detection schemes.

1.4 Machine Learning Approaches for Jamming Detection

The fundamental limitations of traditional, rule-based detection methods—their rigidity, inability to adapt to dynamic conditions, and susceptibility to high false alarm rates—have created a clear and unavoidable imperative to adopt more sophisticated techniques. ML has emerged as the leading paradigm for addressing the complexity of modern wireless security challenges, offering the potential to learn and identify intricate patterns of malicious behavior from high-dimensional network data. This section will first articulate the rationale for applying ML to jamming detection, then provide a critical review of the current state-of-the-art, and culminate in a detailed identification of the significant research gaps that persist in the literature, thereby motivating the novel contributions of this thesis.

1.4.1 The Rationale for Machine Learning in Wireless Security

The core value proposition of Machine Learning (ML) in the context of jamming detection is its ability to move beyond simple, one-dimensional thresholds and instead learn a complex, multi-dimensional decision boundary that separates normal network behavior from anomalous, jammed states. In contrast to traditional methods that require arduous and error-prone manual tuning, ML algorithms can automatically weigh and combine a large set of features, or metrics, from various network layers to build a more robust and adaptive detection model.

The 5G and IoT environment is characterized by extreme complexity and dynamism. Factors such as device mobility, fluctuating channel conditions, and varying network traffic loads create a constantly shifting baseline of "normal" behavior [109]. It is practically impossible for a human operator to define a set of static rules that can accurately capture this complexity and reliably distinguish it from the subtle footprint of an advanced jammer. ML models, however, are designed to learn these complex, non-linear relationships directly from data [95]. By training on examples of both normal and jammed network traffic, these models can discover the subtle, correlated patterns across multiple KPIs that are indicative of a jamming attack, leading to higher detection accuracy and a greater resilience to the environmental noise that confounds

traditional methods [156]. This ability to automate the discovery of complex patterns makes ML an essential tool for building the next generation of intelligent and adaptive security systems.

1.4.2 A Critical Review of Current ML-Based Detection Systems

The application of ML to wireless security is an active area of research, with a variety of techniques being explored for jamming detection and classification. The literature demonstrates the use of supervised, unsupervised, and reinforcement learning methods to address challenges ranging from authentication and access control to malware and jamming detection.

In the domain of jamming detection, a significant body of work has focused on applying and comparing various supervised learning algorithms. For example, studies have used classifiers like Random Forests, Support Vector Machine (SVMs), and Decision Trees to distinguish between different types of jammers, such as constant, random, and reactive [71, 21, 74, 88, 111, 19]. While these models have shown success in classifying distinct jammer types like constant jammers, the results indicate that differentiating between more subtle and similar strategies, such as random versus reactive jammers, remains a significant challenge.

More recently, researchers have begun to explore the use of deep learning for on-device detection. Architectures like Convolutional Neural Networks (CNN) have been implemented on edge devices to perform real-time classification of jamming attacks. These approaches have demonstrated high classification accuracy, often exceeding 90%, but typically within highly controlled laboratory testbed environments [101, 36, 209, 38]. The underlying principle of these advanced systems remains consistent: to leverage a high-dimensional feature space, composed of multiple cross-layer network KPIs, to learn a sophisticated decision boundary that is more resilient and accurate than any single-metric thresholding approach [67].

Despite these promising results, a deeper, critical analysis of the methodologies employed in the existing literature reveals a set of recurring and significant shortcomings. These limitations, which span from the data used for training to the architectural assumptions made for deployment, collectively create a significant "reality gap" between the performance reported in academic papers and the likely efficacy of these systems in real-world, operational environments. The convenience of laboratory setups and the availability of data have, in many cases, shaped the research methodology more than the practical requirements of the problem itself. This has led to the development of models that perform exceptionally well on clean, curated benchmark datasets but are likely too brittle, inefficient, or architecturally unsuitable for deployment in the chaotic and resource-constrained environments where they are most needed.

1.4.3 Identified Research Gaps in Existing Literature

A thorough review of the state-of-the-art in ML-based jamming detection reveals four critical and interconnected research gaps. These gaps are not minor limitations but fundamental issues that currently prevent the widespread, effective deployment of these techniques in practice. This

thesis is explicitly designed to address these shortcomings.

1. **Over-reliance on Simulated Data.** A vast majority of existing studies develop and evaluate their ML models using data generated from network simulators like ns-3 [71, 74, 19, 21, 149, 88]. While simulators are invaluable tools for protocol design and initial testing, they inherently fail to capture the full complexity, unpredictability, and non-idealities of real-world wireless environments. Factors such as multi-path fading, unpredictable interference from other devices, and the subtle hardware-specific characteristics of radio transceivers are often abstracted away or modeled imperfectly. Models trained exclusively on this sanitized, synthetic data are likely to be poorly generalized. When deployed in a real network, they may exhibit a significant drop in performance because they have not been trained to be robust to the noise and variability of the physical world. This limits the credibility and practical relevance of their reported findings.
2. **Limited and Unimodal Feature Sets.** Even among studies that utilize real, over-the-air data, a common limitation is the reliance on a narrow and unimodal set of features. The most frequently used features are physical layer metrics, primarily RSSI and SNR, which are easily accessible from standard wireless drivers [74, 36, 19, 21]. While these metrics are certainly affected by jamming, they do not provide a complete picture of the network's state. An advanced attacker can employ strategies that have a subtle effect on RSSI but a dramatic impact on higher-layer performance, such as PDR or latency. Relying solely on a single modality (e.g., physical layer data) prevents the system from forming a holistic view and makes it susceptible to misdetections, especially if an attacker deliberately masks their activity in one domain while leaving subtle traces in another.⁴⁷ A truly robust system requires a cross-layer approach, integrating features from the physical, MAC, and even application layers, yet such approaches remain rare in the literature.
3. **Centralized Processing and Lack of Scalability.** The predominant architectural model assumed in the literature is centralized. In this model, raw data or features are collected from distributed network nodes and transmitted to a central server or base station where the ML model is executed [36, 209, 38, 208]. This architecture is fundamentally unsuitable for the massive-scale, real-time demands of IoT networks. It introduces significant communication latency, which is unacceptable for detecting and responding to time-critical security threats. Moreover, it places a heavy burden on the network's bandwidth and the energy resources of the battery-powered end devices, which must constantly transmit monitoring data. This approach is not scalable, efficient, or feasible for the intended deployment environment at the network edge [101].
4. **Simplistic and Non-Generalizable Threat Models.** The experimental evaluations in many papers are conducted against overly simplistic threat models. The jamming attacks are often limited to a basic constant or reactive jammer operating under static, controlled

laboratory conditions [21, 25, 71, 88, 100, 149]. The resulting ML models, while effective against these specific, well-behaved threats, are unlikely to be generalizable or robust against the advanced, adaptive, and power-modulated jammers that represent the true state-of-the-art for adversaries. This focus on easily reproducible but unrealistic threats further widens the gap between academic results and real-world security.

These four gaps are deeply intertwined and create a vicious cycle that stifles progress in the field. The scarcity of high-quality, realistic, and publicly available datasets for jamming attacks forces researchers to rely on simulations [193]. The models trained on this unrealistic data are not robust enough for real-world deployment. Because the models are not widely deployed, there is no mechanism for the large-scale collection of real-world attack data that could be used to create better public datasets, thus perpetuating the cycle. Breaking this cycle is the most critical challenge for advancing the field of ML-based wireless security. It requires a methodological shift away from convenient simulations and towards frameworks that can operate in the real world, learn from limited and imperfect data, and address the foundational problem of data scarcity head-on.

1.5 Towards Next-Generation Detection: On-Device and Collaborative Intelligence

The identified gaps in the existing literature—the over-reliance on simulation, the use of simplistic feature sets, and the assumption of centralized processing—point towards the need for a new architectural paradigm for jamming detection. This new paradigm must be grounded in the realities of the 5G and IoT ecosystem: it must be scalable to billions of devices, operate in real-time at the network edge, and function effectively despite the inherent resource constraints of the hardware [52]. This section introduces the forward-looking concepts that form the pillars of this new paradigm and the foundation of this thesis’s contributions. It makes the case for a synergistic architecture that combines on-device intelligence with collaborative sensing and addresses the foundational challenge of data scarcity through the critical evaluation of public datasets and the exploration of advanced synthetic data generation techniques.

1.5.1 The Imperative for On-Device Detection at the Network Edge

The limitations of the centralized processing model make a compelling case for pushing detection capabilities to the network edge, directly onto the end devices themselves. An on-device detection mechanism offers three critical advantages over a centralized one: low latency, scalability, and efficiency [216]. By performing analysis locally, an on-device system can detect a threat in real-time without the delay incurred by transmitting data to a central server and waiting for a response. This is crucial for time-sensitive applications where immediate action is required. Furthermore, as the detection logic resides on each device, the system scales naturally as the

network grows, avoiding the communication and processing bottlenecks that would cripple a centralized server in a massive IoT deployment [175]. Finally, it is far more efficient in terms of bandwidth and energy, as raw, high-volume sensor data does not need to be constantly transmitted over the wireless network; only low-volume alerts or summary statistics are communicated when necessary [101].

The primary challenge to implementing on-device detection has always been the severe resource constraints of IoT hardware. However, the emerging field of TinyML [63] is dedicated to overcoming this challenge by developing techniques to design and deploy high-performance ML models on low-power microcontrollers and other embedded systems. This involves model compression, quantization, and the design of resource-aware neural network architectures. By leveraging TinyML principles, it becomes feasible to create lightweight yet powerful ML models for jamming detection that can operate within the strict memory, computational, and energy budgets of typical IoT devices [131]. This shift towards on-device intelligence is a necessary step to create a security solution that is architecturally compatible with the distributed nature of the IoT.

1.5.2 The Power of Collaborative Sensing in Distributed Networks

While on-device detection provides the low-latency, scalable foundation for a next-generation security architecture, it is not a complete solution on its own. A single device has an inherently limited, localized perspective of the network. It can observe its own performance metrics, but it cannot easily distinguish between a localized issue (e.g., a nearby source of interference affecting only itself) and a coordinated, large-scale attack affecting a wide area. This ambiguity can lead to false positives or an inability to grasp the full scope of a sophisticated threat.

To overcome this limitation, the concept of collaborative or distributed sensing, which has been extensively explored in the context of Wireless Sensor Networks (WSN), can be integrated with on-device intelligence [186, 208, 125]. In a collaborative framework, individual nodes do not operate in isolation. Instead, they share their local findings, specifically their high-level, on-device detection results, with their neighbors or with a local coordinator. This sharing of information allows the network to build a more accurate, collective, and context-rich understanding of its state [60]. For example, if a single node detects a drop in its PDR, it might flag a potential anomaly. However, if it then learns through collaboration that all of its neighbors are experiencing a similar drop in PDR simultaneously, the confidence that this is a widespread jamming attack, rather than an individual hardware failure, increases dramatically.

Prior work in WSN has proposed various collaborative mechanisms, ranging from centralized schemes where nodes report their status to a cluster head [141, 96, 140], to fully distributed approaches where nodes directly exchange performance metric information with their immediate neighbors to cooperatively identify and isolate jammed zones [60, 54]. This combined knowledge enables the network to defend itself more efficiently and accurately.

On-device intelligence and collaborative sensing are therefore not mutually exclusive solu-

tions but are two synergistic components of an optimal detection architecture. The most effective model is a hierarchical one: raw data processing and initial, lightweight anomaly detection are performed on-device to ensure low latency and scalability. Then, the high-level outputs of these local models are shared collaboratively among nodes to build a robust, system-wide consensus. This architecture achieves the best of both worlds: it retains the real-time responsiveness of edge processing while gaining the global context and accuracy of a distributed system, all while minimizing the communication overhead that plagues traditional centralized approaches. This synergistic model forms the architectural blueprint for the contributions of this thesis.

1.5.3 Addressing Data Scarcity: A Review of Public Datasets and Synthetic Data Generation

The development and validation of any ML model, whether it operates on-device or collaboratively, is fundamentally dependent on the availability of high-quality, relevant data. In the domain of network security, and particularly for physical layer attacks like jamming, the scarcity of suitable public datasets is one of the most significant impediments to research and development [193]. This scarcity forces a reliance on unrealistic simulations and creates the "reality gap" discussed previously. Addressing this data scarcity problem is therefore a critical prerequisite for building robust and reliable detection systems. This requires a critical evaluation of existing public datasets and an exploration of advanced data augmentation techniques to supplement them.

Public Datasets and Their Limitations: Several public datasets have been created for the evaluation of Network Intrusion Detection Systems (NIDS), with UNSW-NB15 [163] being a prominent example often cited as an improvement over older datasets like KDD99 [162, 57]. The UNSW-NB15 dataset was generated in a lab environment and contains a hybrid of real normal network traffic and synthetically generated contemporary attack behaviors, including nine attack types such as DoS, Exploits, and Worms. While valuable for general NIDS research, this and other similar public datasets suffer from a number of critical limitations when considered for the specific task of modern jamming detection:

- **Timeliness and Static Nature:** Network traffic patterns and attack techniques evolve constantly. Public datasets are static snapshots in time and quickly become outdated, failing to represent the current threat landscape [79].
- **Lack of Realism:** Most public datasets are generated in emulated laboratory testbeds. The resulting traffic often lacks the complexity, diversity, and noise of real-world enterprise or IoT networks, and the attack scenarios may not be realistic or effective.
- **Class Imbalance:** In real networks, attack traffic is rare compared to benign traffic. This is reflected in datasets, which are often highly imbalanced. This class imbalance can

cause ML models to become biased towards the majority (normal) class, resulting in poor detection rates for the minority (attack) class [196].

- **Narrow Focus:** Datasets like UNSW-NB15 are focused on higher-layer network intrusions (e.g., exploits, shellcode). They do not contain the physical layer data (e.g., detailed RF signal characteristics) necessary to train and evaluate models for detecting physical layer attacks like RF jamming.

Data Augmentation Techniques: Given the limitations of existing datasets, researchers turn to data augmentation techniques to create additional training data, particularly for the under-represented attack class. This shifts the problem from one of "Big Data" processing to one of "Smart Data" generation, which is the intelligent creation of relevant data where none exists.

- **Synthetic Minority Over-sampling Technique (SMOTE):** SMOTE [42] is a widely used and straightforward data augmentation technique for addressing class imbalance. Instead of simply duplicating existing minority class samples (which can lead to overfitting), SMOTE generates new, synthetic samples by interpolating between a minority class instance and one of its randomly chosen nearest neighbors in the feature space [235]. While effective in many scenarios, SMOTE has notable limitations. It operates without regard to the majority class, which can lead to the creation of noisy samples in overlapping class regions. If the minority class is very sparse or scattered, the linear interpolation may produce unrealistic samples that do not conform to the true underlying data distribution, a problem known as overgeneralization [27]. Furthermore, SMOTE is primarily designed for continuous, tabular data and is less effective for complex, high-dimensional data or time-series data where temporal dependencies are critical.
- **GANs:** GANs represent a more sophisticated and powerful approach to synthetic data generation [80]. A GAN consists of two neural networks, a Generator and a Discriminator, which are trained in an adversarial contest. The Generator's goal is to create synthetic data samples that are indistinguishable from real data, while the Discriminator's goal is to correctly identify whether a given sample is real or synthetic (generated). Through this competitive training process, the Generator learns to capture the underlying probability distribution of the training data and can produce high-fidelity, diverse synthetic samples [213]. In the context of cybersecurity, GANs have been successfully used to generate realistic synthetic attack samples to balance datasets, thereby improving the performance of Intrusion Detection Systems (IDS), particularly for rare attack types [135]. While GANs can overcome many of SMOTE's limitations by learning complex data distributions, they are also more computationally expensive, can be difficult to train stably, and may suffer from issues like mode collapse, where the generator produces a limited variety of samples [17].

The future of ML in wireless security, therefore, lies not in passively waiting for the perfect "Big Data" set to appear, but in actively and intelligently augmenting small, high-quality, real-world datasets with synthetically generated "Smart Data." The exploration of advanced generative models like GAN is not merely an optional preprocessing step but a critical research direction for creating ML systems that are robust enough to handle the rare and novel threats characteristic of the security domain.

1.6 Summary and Thesis Motivation

This chapter has provided a comprehensive review of the background and related work essential to understanding the security challenges in modern 5G and IoT systems, with a specific focus on the threat of RF jamming. The analysis has progressed from a high-level architectural overview to a detailed critique of existing detection methodologies, culminating in the identification of critical research gaps that motivate the work of this thesis.

The review began by establishing the complex problem space. The architectural convergence of software-defined 5G networks and the massively scaled, resource-constrained IoT has created an unprecedentedly large and heterogeneous attack surface. While vulnerabilities exist at every layer, the open and shared nature of the wireless medium makes the physical layer a universal point of weakness. The threat of RF jamming has evolved significantly, moving beyond simple, brute-force DoS attacks to highly sophisticated, intelligent, and adaptive strategies. These advanced jammers are energy-efficient, stealthy, and in some cases, are explicitly designed to be anti-forensic, targeting the very countermeasures deployed to detect and locate them.

The survey of existing solutions revealed that current defense paradigms are ill-equipped to handle this evolving threat. Signature-based detection is, by definition, ineffective against novel attacks. While the anomaly-based detection paradigm is necessary, traditional implementations that rely on static, manually-tuned thresholds for network metrics are brittle, inflexible, and prone to high false alarm rates in dynamic wireless environments. The application of ML has shown significant promise in overcoming these limitations. However, a critical analysis of the current ML literature has exposed a significant "reality gap." The field is characterized by an over-reliance on unrealistic simulated data, the use of limited and unimodal feature sets that fail to capture a holistic network view, and the prevalence of centralized processing architectures that are not scalable or efficient enough for real-time deployment at the network edge. This has resulted in a vicious cycle, where a lack of realistic, publicly available data stifles the development of robust, deployable models.

This thesis directly confronts and addresses these identified research gaps. It aims to bridge the reality gap by moving beyond simulation to develop and characterize a system for detecting anomalous behavior using real-world data collected from on-device sensors. It breaks from the unimodal and centralized paradigms that dominate the literature by proposing a novel, synergistic framework that combines:

1. **Lightweight, On-Device Anomaly Detection:** By leveraging the principles of TinyML, this research will develop and implement ML models that are efficient enough to perform real-time jamming detection directly on edge devices. This provides the immediate, localized context necessary for a rapid response.
2. **Enhanced Model Robustness through Generative Data Augmentation:** Recognizing that the performance of on-device models is fundamentally limited by the availability of diverse training data, this thesis introduces a tailored GAN-based approach to create high-fidelity synthetic time-series data that captures the complex characteristics of both normal and anomalous wireless behavior.
3. **Collaborative Intelligence:** This research will design and evaluate a framework that allows these edge devices to share their high-level, local findings. This collaborative process builds a distributed, system-wide consensus that enhances detection accuracy, provides global context to distinguish widespread attacks from local faults, and significantly reduces the false positives that plague isolated anomaly detection systems.

By designing, implementing, and rigorously evaluating this comprehensive framework, this thesis aims to make a significant contribution to the state-of-the-art. It seeks to provide a scalable, efficient, and robust solution for detecting advanced jamming attacks, thereby enhancing the security and reliability of critical 5G and IoT systems.

On-Device Jamming Detection using Native System Logs

The preceding chapter reviewed the landscape of jamming attacks in 5G networks and surveyed existing detection methodologies. This review highlighted several critical gaps: traditional hardware-based solutions lack scalability for widespread deployment, while many current ML approaches suffer from a reliance on simplistic simulations and a lack of rigorous validation against sophisticated, real-world threats. Furthermore, the identified limitations of static public datasets underscore the need for a more dynamic and realistic approach to model development and evaluation.

To address these shortcomings, this chapter presents the first primary contribution of this thesis: a scalable and holistic framework for on-device jamming detection that does not rely on specialized hardware. It details the SHIELD framework, a complete pipeline for collecting and processing native system logs to train and, crucially, to rigorously evaluate the robustness of ML classifiers. The objective is to establish the fundamental feasibility of performing accurate and resilient jamming detection directly on COTS devices, thereby providing the foundational building block for the work in subsequent chapters.

2.1 Motivation: A Scalable, Hardware-Agnostic Approach

The existing approach to detecting and mitigating radio jamming attacks has been fundamentally constrained by a reliance on specialized, external hardware. Techniques centered on spectrum analyzers and advanced signal processing systems, while reliably effective, are characterized by prohibitive costs and a significant lack of scalability. This section explains these limitations and their consequent impracticality for the widespread, ubiquitous deployment required to secure the vast ecosystems of 5G and the IoT.

2.1.1 Foundational Principles of Hardware-Based RF Fingerprinting

Radio Frequency Fingerprinting (RFF) is a physical-layer identification technique that enables a receiver to authenticate a transmitter without relying on cryptographic protocols [108, 85]. The core principle is that every radio transmitter possesses a unique, inherent signature embedded within its emitted signal, a direct result of minute imperfections in its analog hardware components [199]. These imperfections are unintentional byproducts of the manufacturing process, where microscopic variations in materials and fabrication lead to unique characteristics in components like power amplifiers, digital-to-analog converters, oscillators, and frequency mixers [221].

These hardware variations manifest as subtle but distinctive distortions in the transmitted signal, such as carrier frequency offset, phase noise, I/Q imbalance, and power amplifier non-linearity [199]. Collectively, these traits constitute a device's fingerprint, which is inherently difficult to mask or replicate, making it a robust tool for identifying a specific piece of hardware [108].

The analysis to extract this fingerprint can be performed on two distinct parts of a signal transmission. Transient-based RFF focuses on the initial power-on phase of a signal, capturing the unique characteristics as the transmitter transitions from an off to an on state [199]. This transient period is often short-lived and rich in device-specific features but typically requires high sampling rates for accurate capture. In contrast, steady-state RFF analyzes the modulated portion of the signal, where the data is being transmitted, leveraging features embedded during the stable operation of the transmitter.

For an RFF to be effective as a security primitive, it must ideally possess several key properties. These include: universality, meaning every device has the feature; uniqueness, ensuring no two devices share the same fingerprint; permanence, indicating the fingerprint is time-invariant; collectability, allowing for quantitative measurement with available equipment; and robustness, meaning it remains stable against environmental factors and device-related aspects like temperature and voltage fluctuations [221]. While this represents the theoretical ideal, the practical challenges in achieving these properties form a significant part of the limitations discussed in subsequent sections.

2.1.2 The Scalability and Economic Impasse in Massive Deployments

The traditional, centralized paradigm for RFF-based security faces an insurmountable wall of scalability and economic unsustainability. The sheer scale of 5G and IoT networks, which are projected to encompass tens of billions of devices, renders architectures that rely on centralized processing and expensive, specialized hardware fundamentally non-viable. This section argues that traditional RFF deployment models are caught in an unbreakable trilemma of scalability, security, and cost, creating a strategic dead end that necessitates a complete architectural rethinking.

The Computational Bottleneck of Centralized Intelligence

The prevailing model for ML-based RFF and jamming detection involves collecting raw signal data (e.g., I/Q samples) from distributed sensors and forwarding it to a powerful, centralized server or cloud platform for processing and analysis [177]. While this approach simplifies the design of the end-point sensors, it creates an unsustainable bottleneck at the network core. The vast amount of data generated by potentially billions of IoT devices would place an immense and continuous strain on both network bandwidth and computational resources. This model is simply not built for the real-time demands of security monitoring at scale.

The centralized architecture also introduces unacceptable latency. The time required for data to travel from the network edge to a central cloud, be processed by a complex ML model, and for a response or alert to be sent back is far too long for many critical applications [50]. For 5G URLLC services, which operate on latency budgets of less than one millisecond, a cloud-based detection loop is an absolute non-starter [94].

Furthermore, the aggregation of raw signal data from countless devices in a single location creates profound data privacy concerns [94]. As established in Section 2.1.1, RFF can be used to track and profile users. A centralized repository of this data would become a high-value target for attackers, and its existence raises significant regulatory and ethical questions.

An Economic Analysis: The Asymmetry of Defense Costs vs. Attack Capabilities

Beyond the technical and operational hurdles, there is a stark economic reality that renders traditional hardware-based defense unsustainable. A severe and unfavorable asymmetry exists between the cost of launching a sophisticated jamming attack and the cost of deploying a high-fidelity system to detect it.

Jamming and spoofing attacks can be implemented using low-cost, commercially available hardware. SDR, which can be programmed to generate a wide variety of malicious signals, are readily accessible and can be purchased for a few hundred dollars. This low barrier to entry means that an adversary can affordably deploy a large number of attack devices.

In sharp contrast, the equipment required for high-precision RF analysis, which is necessary to capture the subtle features needed for reliable fingerprinting, is orders of magnitude more expensive. High-end laboratory equipment such as vector signal analyzers, gigasample-rate oscilloscopes, and high-frequency spectrum analyzers can cost tens or even hundreds of thousands of dollars per unit [199]. While more affordable handheld analyzers and high-end SDR exist, they still represent a significant cost when considering deployment across the vast geographical areas covered by 5G and IoT networks. The following table provides a quantitative comparison that illustrates this stark economic disparity (data synthesized from sources [224]).

This economic imbalance creates an untenable situation for network defenders. An attacker can field a distributed network of dozens or hundreds of low-cost jammers for the price of a single high-end defensive sensor. Attempting to scale a defense based on expensive, specialized

Equipment Type	Primary Use Case	Frequency Range	Price Range
High-End Spectrum/Signal Analyzer	Defensive Sensing/Analysis	2 Hz – 50 GHz	\$40,000 – \$180,000+
Benchtop/Basic Spectrum Analyzer	Defensive Sensing/Analysis	9 kHz – 20 GHz	\$14,000 – \$29,000
Handheld Spectrum Analyzer	Defensive Sensing/Field Analysis	100 kHz – 7.1 GHz	\$350 – \$7,000
High-Performance SDR (e.g., USRP)	Defensive Sensing/Research	up to 6 GHz	\$1,500 – \$8,000
Low-Cost SDR (e.g., HackRF, LimeSDR)	Offensive Jamming/Spoofing	1 MHz – 6 GHz	\$150 – \$500

Table 2.1: Comparative Cost Analysis of RF Sensing and Attack Hardware

hardware is, therefore, an economically losing proposition.

This analysis leads to a critical conclusion: the traditional RFF paradigm is constrained by a scalability-security-cost trilemma. A network can be designed to achieve any two of the following three attributes, but not all three simultaneously:

- **Massive Scalability:** Supporting billions of dynamic devices.
- **Robust Security:** Providing high-accuracy, real-time threat detection.
- **Low Cost:** Ensuring economic viability for widespread deployment.

To achieve robust security at massive scale, one would need ubiquitous deployment of expensive sensors and massive centralized computing power, violating the low-cost constraint. To achieve massive scale at low cost, one would have to use cheaper sensors and infrequent model updates, violating the robust security constraint. This trilemma demonstrates that the centralized, hardware-intensive approach is a strategic dead end. The only path forward is to fundamentally change the architectural paradigm by decentralizing intelligence and leveraging the vast network of existing commodity hardware.

To address this challenge, this chapter advocates for a fundamental paradigm shift: moving from hardware-dependent solutions to a scalable, cost-effective, and software-centric approach. The core motivation of this chapter is to establish the feasibility of performing robust jamming detection by leveraging data sources that are natively available on COTS devices. This work posits that the rich telemetry and operational metrics embedded within system logs—generated by both UE and the gNB during routine network operations—can serve as effective indicators of anomalous behavior indicative of jamming.

By utilizing these pre-existing data streams, the detection system can be deployed directly on the network entities themselves, creating a self-contained and easily distributable security

solution. This hardware-agnostic methodology eliminates the need for additional probes or infrastructure, thereby offering a practical and robust solution for securing large-scale 5G deployments.

2.2 The SHIELD Framework

To address the challenges of scalability and cost, this work introduces SHIELD. This framework is presented as an end-to-end solution that integrates a controllable wireless jamming module with a comprehensive ML pipeline, designed specifically for assessing jamming interference using data from both the UE and the gNB. The framework is engineered to be holistic by capturing data from both ends of the communication link and scalable by eschewing dependencies on external hardware.

The architecture of the SHIELD framework, depicted in Figure 2.1, is not merely a linear pipeline but a self-contained, closed-loop system for rigorous scientific validation. It is composed of four principal components that form a complete workflow from threat emulation to performance evaluation:

1. **Controlled Jamming Interference:** This component implements a realistic and challenging threat model, specifically a power-modulated jammer. Its purpose is to create a controlled, repeatable, and sophisticated interference environment that serves as a reliable ground truth for evaluating the efficacy of detection defenses. This moves beyond the simplistic jammer profiles often used in prior work.
2. **Cross-End Data Collection Mechanism:** SHIELD employs a mechanism to synchronously capture native logs and traces from both endpoints of the 5G communication link: the UE and the gNB. This dual-source approach provides a comprehensive, multi-layer, and end-to-end view of the network's behavior under jamming, critically avoiding any reliance on costly external probes or specialized hardware.
3. **Robust Preprocessing Methodology:** Recognizing that raw log data is often noisy, asynchronous, and sparse, the framework includes a multi-stage preprocessing pipeline. This component is responsible for parsing, cleaning, interpolating, and transforming the raw data into a feature-rich, synchronous format that is optimized for machine learning analysis.
4. **Rigorous ML Assessment:** To combat the issue of overly optimistic performance estimates common in the literature, SHIELD incorporates a novel two-data-collection strategy. It utilizes a primary dataset for model training and validation, and a distinct, separately collected robustness test set to ensure that the resulting models are genuinely resilient to real-world environmental variations and not merely overfitted to a static set of lab conditions.

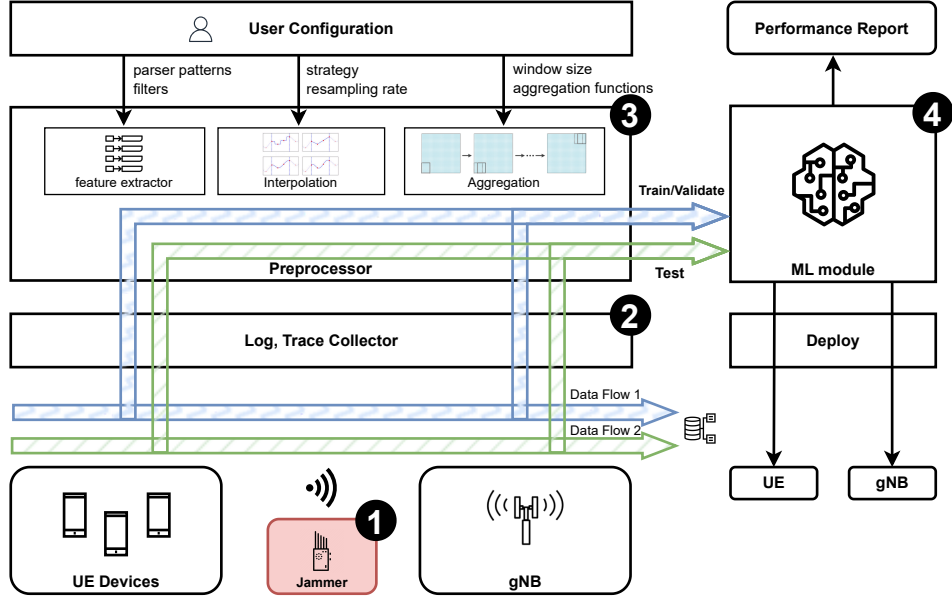


Figure 2.1: Architecture of the SHIELD Framework.

A core tenet of this framework is a rigorous methodology for evaluating model robustness, which addresses a common flaw in many data-driven studies. Our observation, particularly from the WSN domain [254, 74], is that data collected in a single, continuous session tends to be homogeneously distributed, regardless of its duration. Consequently, a standard train-test split on such a dataset inevitably leads to information leakage, where the feature distributions of the training and testing sets are nearly identical. This explains why many models reported in the literature achieve near-perfect performance in testing but fail to generalize in practice.

To overcome this, our approach mandates a multi-data collection strategy. By collecting a "primary dataset" for training and an entirely separate robustness test set under demonstrably different network conditions, we ensure the creation of two datasets with distinct statistical properties. This structure provides a true and stringent test of a model's generalizability, forcing an evaluation of its performance against genuinely unseen data. The value of SHIELD, therefore, lies not only in the final detection model it produces but in the scientifically sound process it defines for validating models in a way that is directly applicable to real-world deployment.

2.3 Experimental Testbed and Threat Model

To validate the SHIELD framework and to overcome the limitations inherent in purely simulated environments, a quasi-realistic 5G testbed was constructed, enabling the emulation of sophisticated jamming attacks and the collection of high-fidelity data from both network infrastructure

and a commercial end-user device.

2.3.1 Testbed Components

The experimental environment was established in a non-anechoic indoor setting, reflecting a typical deployment scenario with realistic ambient noise and signal propagation effects. The testbed integrated open-source software with a combination of commercial and SDR hardware to create a functional 5G network.

- **CN and RAN:** The 5G Core (5GC) was implemented using the open-source Open5GS project [173]. The RAN was realized with the software-emulated srsRAN stack [225].
- **Hardware Components:** The physical hardware components and their arrangement are depicted in Figure 2.2a. Two Nuand bladeRF 2.0 micro SDRs served as the physical layer transceivers. One SDR was configured to operate as the gNB, while the second was dedicated to acting as the jammer. A COTS OnePlus Nord 2T 5G smartphone was used as the UE, ensuring that the data collected from the device side accurately reflects the behavior and logging mechanisms of a real-world consumer product.
- **Software and Configuration:** Logs from the Android-based UE were collected in real-time using the Android Debug Bridge (ADB) utility. Logs from the gNB were collected directly from runtime. During experiments, network traffic was simulated using the iPerf3 utility, and the jammer's efficacy was confirmed by observing its significant negative impact on KPIs, including network latency, jitter, and bandwidth.
- **Physical Setup:** The physical environment was deliberately altered between collection sessions, as shown in Figure 2.2b. The UE was positioned at a distance ranging from 10 to 15 meters from the gNB, with and without obstacles to generate distinct datasets for robustness testing. The jammer was placed in close proximity to the gNB, approximately 1 meter away, to effectively interfere with the communication link. The jammer SDR was programmed to operate specifically in the 5G n2 uplink frequency band (1850-1910MHz).

2.3.2 Threat Model

The threat model employed within the SHIELD framework was deliberately designed to be more subtle and challenging than those commonly found in the literature.

Adversary Objective: The adversary's primary goal is not to execute a brute-force DoS attack that would completely sever connectivity and trigger obvious network alarms. Instead, the objective is to achieve a subtle degradation of performance, prolonging the period of interference while evading immediate detection. This stealthy approach poses a greater challenge to conventional monitoring systems.

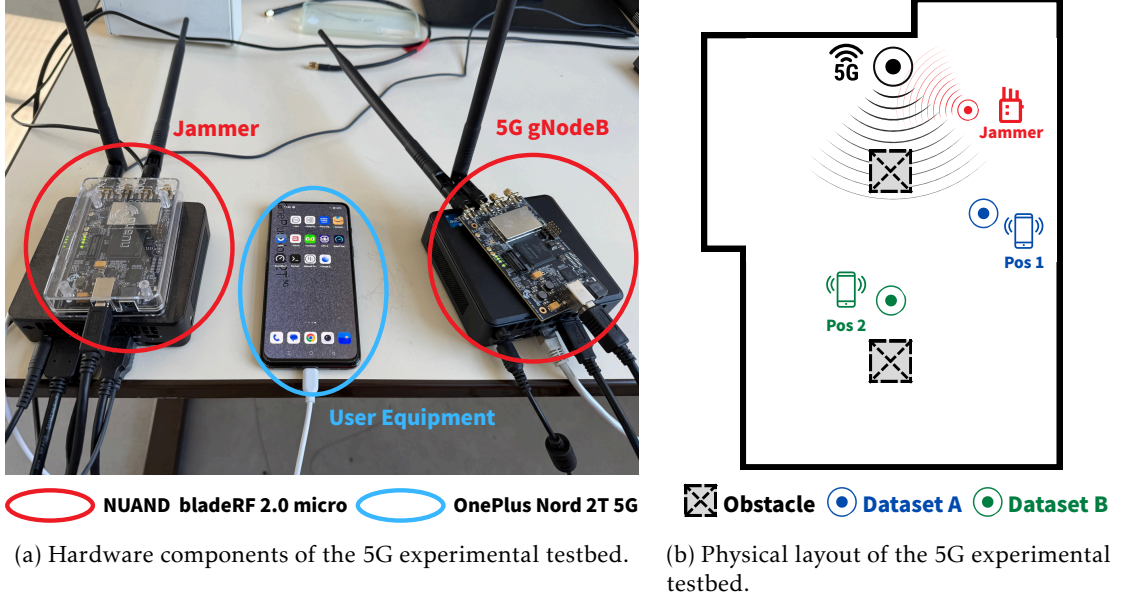


Figure 2.2: Experimental setup for data collection, illustrating the different positions and conditions for the primary dataset (A) and the robustness test set (B).

Technical Implementation: To realize this objective, a power-modulated jammer was implemented. This design marks a significant departure from simplistic constant-power or reactive jammers. By leveraging sinusoidal amplitude patterns, the jammer generates a variable-power noise profile, resulting in unpredictable and dynamic interference. This time-varying nature is specifically intended to defeat detection techniques that rely on simple statistical thresholds for metrics like RSSI or SNR.

Operational Parameters: The behavior of the power-modulated jammer is governed by four key, configurable parameters, which allow for precise and diverse control over the characteristics of the interference signal:

- Sampling Rate r_{smp} : Controls the frequency resolution of the signal, enabling more precise manipulations of jamming characteristics as the rate increases.
- Slot Length L_{slot} : Defines the interval, measured in samples, at which the amplitude pattern resets. It allows the jammer to generate structured, periodic bursts of interference, mimicking pulsed or intermittent attack scenarios.
- Amplitude Range (min_A, max_A): Governs the power of the emitted signal, supporting a wide variety of jamming strengths.
- Pattern Frequency f_{pat} : Determines the rate of sinusoidal oscillations in the signal's amplitude, allowing the jammer to vary its power in real time and evade straightforward

detection.

The formal logic for the generation of the jamming signal is detailed in Algorithm 1, which structures the interference into slots to create dynamic bursts that challenge the static noise filtering capabilities of wireless systems. The efficacy of this attack was confirmed by observing its impact on network performance indicators, which showed (in Figure 2.3) a clear decrease in bandwidth and an increase in jitter and packet loss upon jammer activation.

Algorithm 1 Power-Modulated Jamming.

Require: Sampling Rate r_{smp} , Slot Length L_{slot} , Amplitude Range (min_A, max_A) , Pattern Frequency f_{pat}

Ensure: Jamming Signal J_{signal}

- 1: Initialize time index $n \leftarrow 0$
- 2: Compute amplitude modulation factor: $\mathcal{A}_m \leftarrow \frac{1}{2}(max_A - min_A)$
- 3: **while** Jamming is active **do**
- 4: Compute amplitude modulation for current time step:

$$\mathcal{A}(n) \leftarrow min_A + \mathcal{A}_m \times \left(1 + \sin \left(\frac{2\pi n f_{pat}}{r_{smp}} \right) \right)$$

- 5: Generate jamming signal:

$$J_{signal}(n) \leftarrow \mathcal{A}(n) \times Noise(n)$$

- 6: Transmit $J_{signal}(n)$
 - 7: Increment time index: $n \leftarrow n + 1$
 - 8: **if** $n \bmod L_{slot} = 0$ **then**
 - 9: Reset amplitude pattern for the next slot
 - 10: **end if**
 - 11: **end while**
-

The selection of this power-modulated jammer is a strategic choice that functions as a "validity filter" for detection methodologies. Many prior works have focused on simplistic jamming profiles, which can often be detected by monitoring for overt and consistent disruptions. The power-modulated jammer, by design, creates interference that is neither overt nor consistent. Its continuously changing power level means that at any given instant, its signal strength might fall within a range considered "normal," rendering static threshold-based detection on metrics like RSSI ineffective. This threat model was therefore specifically engineered to create a problem that cannot be solved by the naive methods benchmarked from the literature. The ability of the SHIELD pipeline to successfully detect this sophisticated threat is what validates its more complex, pattern-based approach to feature engineering. The threat and the solution are thus co-designed to demonstrate a significant leap in detection robustness.

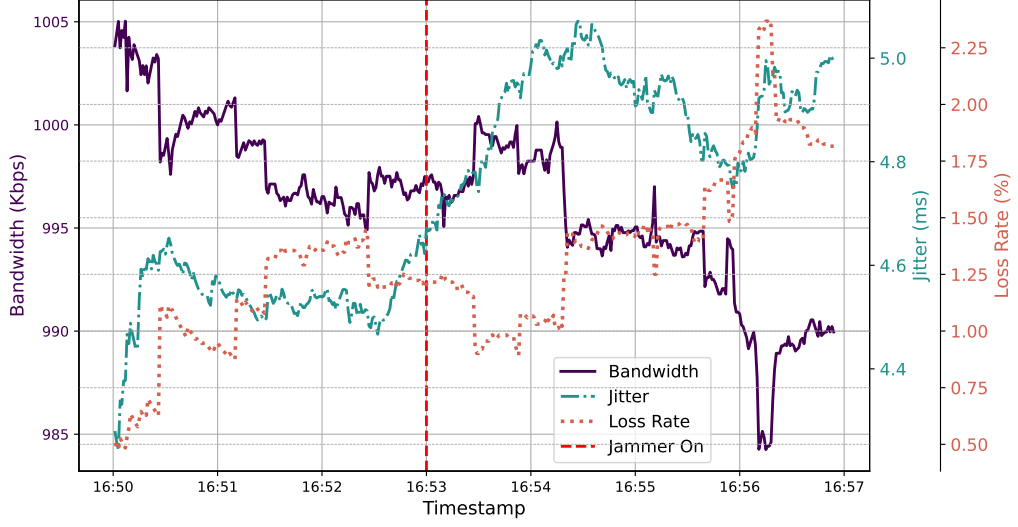


Figure 2.3: Impact of Power-Modulated Jamming on KPIs from iperf3.

2.4 A Holistic Data Collection and Preprocessing Pipeline

A core contribution of the SHIELD framework is its robust and efficient data processing pipeline. This multi-stage methodology is far more than a simple data cleaning procedure. It is a system of strategic information enrichment designed to transform raw, unstructured, and often noisy log data into a high-quality, feature-rich format optimized for ML. The fundamental purpose of this pipeline is to convert low-level, asynchronous log entries into a high-dimensional feature space where the subtle, temporal signatures of sophisticated jamming attacks become mathematically apparent and distinguishable from benign network fluctuations. In this section, we introduce a sequence of carefully designed steps to extract informative features, resolve critical temporal inconsistencies, and construct representative aggregated samples that capture the dynamics of network behavior under interference.

2.4.1 Synchronized Data Acquisition from UE and gNB Logs

The foundational principle of the SHIELD data pipeline is a collection strategy centered on capturing native logs concurrently from both endpoints of the communication link: the UE and the gNB. This dual-source approach is fundamental to achieving a holistic, end-to-end, and multi-layer view of network behavior. This is a critical departure from prior work that often focused on a single network entity.

The practical implementation of this synchronized collection process is illustrated in Figure 2.4. As the diagram shows, after the initial network setup and UE connection authentication, logging processes are initiated in parallel on both the UE and the gNB before traffic generation

begins. On the UE, logs are captured from the Android radio buffer via the ADB, while corresponding traces are simultaneously recorded on the gNB. This synchronized start is crucial for establishing a common temporal baseline, which simplifies the subsequent alignment and fusion of the two disparate data streams.

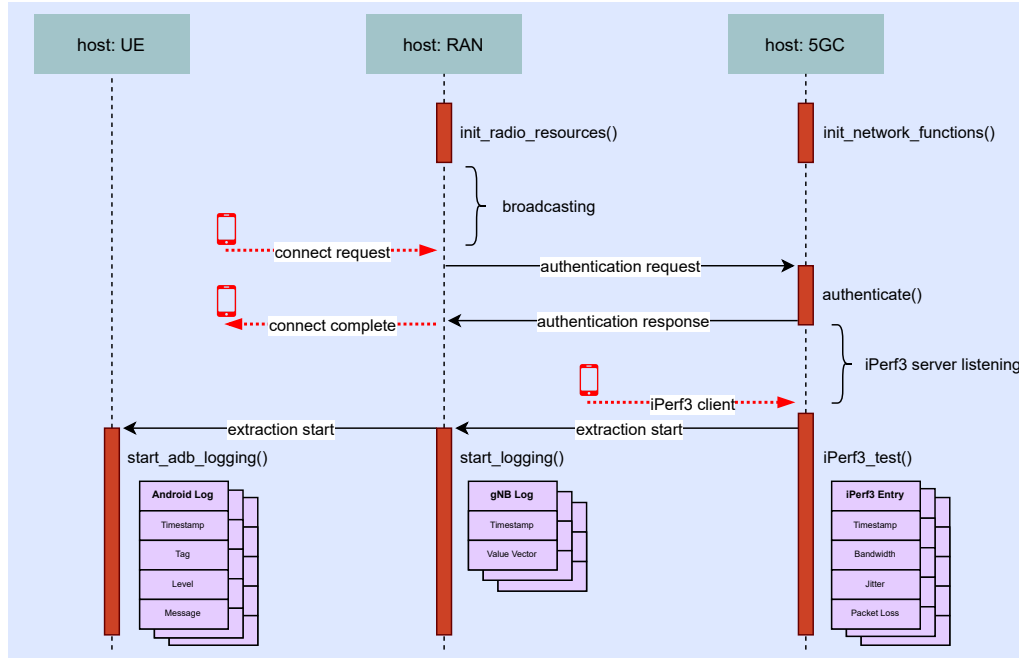


Figure 2.4: End-to-End Synchronized Data Collection Flow in a 5G Infrastructure.

This strategy offers four distinct advantages for robust detection:

- **Improved Data Accessibility:** By leveraging device-level metrics that are available by default in system logs, the framework eliminates any dependency on external probes, spectrum analyzers, or other specialized hardware. This makes the system entirely self-contained and highly scalable for large, distributed deployments.
- **Multi-Layer System View:** The collected data provides a uniquely comprehensive perspective, spanning both low-level Medium Access Control (MAC)-layer metrics and higher-level system insights. This cross-layer visibility is crucial for accurately diagnosing the system-wide impact of an attack.
- **End-to-End Observation:** Monitoring both the UE and gNB provides a complete picture of how jamming interference impacts the entire communication process. This end-to-end perspective is vital for detecting subtle anomalies that might be missed by observing only one side of the link.
- **Real-Time Detection Capability:** Because the system relies solely on existing device logs,

it can function autonomously, enabling on-device, online jamming detection. This makes it feasible to deploy lightweight, self-contained algorithms directly on network hardware.

A key strength of this approach is the richness of the captured data. The native logs are not monolithic, as they contain a diverse array of information that provides a multi-faceted view of the system's state. These can be broadly categorized as follows:

MAC-Layer Metrics: These are direct, quantitative measurements of the physical radio link's performance. They provide an immediate, low-level indication of the connection quality. Examples from the collected logs include:

- From the UE: Signal quality reports including RSSI, Reference Signal Received Power (RSRP), Reference Signal Receiver Quality (RSRQ), and Signal-to-Interference-plus-Noise Ratio (SINR).
- From the gNB: Uplink channel metrics related to Physical Uplink Shared Channel (PUSCH) and Physical Uplink Control Channel (PUCCH).

System-Level Insights: These logs provide higher-level context about the operational state of the device and the network's response to changing conditions. They often reflect the consequences of the MAC-layer behavior. Examples include:

- UE AT command records from the modem, such as ETHERMAL (which provides thermal reports that can indicate device stress from increased processing), ECSQ (logging signal quality requests initiated by the modem), and ERFTX (related to transmit power settings).
- UE Modem activity information extracted from Android Radio Interface Layer (RIL) manager, such as modem sleep/idle time and operating frequency range.
- RAN statistics such as Uplink/Downlink bit rates and packet delivery counts, as well as higher-level scheduling decisions that reflect how the base station is allocating resources in response to channel conditions (Channel Quality Indication (CQI)).

By capturing both categories of data, the framework can build a more complete causal picture of a jamming attack. For instance, a drop in SINR might be accompanied by an increase in ETHERMAL measures, suggesting the modem is working harder to compensate for the poor signal, thereby providing a more robust signature of interference than either metric alone.

Finally, to ensure the resulting dataset was both realistic and comprehensive, data was captured under two distinct traffic profiles. First, the iPerf3 utility was used to generate controlled and replicable traffic. Second, data was collected while running real-world applications, such as live YouTube streaming, on the UE to ensure the dataset reflects diverse usage patterns. It is important to note that while KPIs from the iPerf3 traffic generator (such as bandwidth, jitter, and packet loss) are measured to demonstrate and quantify the jammer's impact on network

performance, they are deliberately excluded from the feature set used for detection. Since these KPIs are the direct targets of an adversary, relying on them for detection would be circular. Instead, the SHIELD framework focuses on the underlying system and radio metrics that reflect the cause of the performance degradation, not the effect itself, leading to a more fundamental and robust detection mechanism.

2.4.2 Feature Engineering from Native Logs

Once the raw data is collected, the next critical stage in the pipeline is feature engineering. This process transforms the unstructured or semi-structured log entries into a clean, quantitative format that is suitable for ML algorithms. The methodology differs significantly between the UE and gNB due to the inherent differences in their log formats.

Parsing UE Logs with Regular Expressions

The logs generated by the UE's Android radio buffer are text-based and semi-structured, often arriving in key-value formats with embedded descriptive tags.

A raw log entry containing signal strength reports, for example, might look like this:

```
mOrigSignalStrength:{... level=0 parametersUseForLevel=1,mNr=
  CellSignalStrengthNr:{ csiRsrp = 2147483647 csiRsrq = 2147483647
  csiCqiTableIndex = 2147483647 csiCqiReport = [] ssRsrp = -88
  ssRsrq = -9 ssSinr = 2 level = 2 parametersUseForLevel = 1 },
  primary=CellSignalStrengthNr}
```

This format is human-readable but not directly ingestible by a ML model. To extract the crucial RSRP measurement (-88 in this case), adaptable regular expression patterns are applied. For the example above, we use the pattern "`(\\w+)\\s?=\\s?([\\^,\\}\\{:\\s]+)`" to match all key-value pairs and capture the integer value associated with the required tags.

AT entries are slightly different:

```
[0] AT< +ECSQ: 72,69,-322,1,1,-29,-351,32768,10,3,32767,32767,32767 (
  RIL_URC_READER, tid:514028494080)
```

This format contains a header message and a sequential value list. We use the pattern "`\\s*[0\\ ] AT< \\+ECSQ:\\s+(\\.+)\\s+(RIL_URC_READER, tid:\\d+\\)`" to extract the integer list and drop the remaining contents.

Similar patterns are designed for each type of message to extract key measurements and contextual signals. This approach is highly flexible. The parsing rules can be customized to match the specific log formats of different device manufacturers or Android OS versions, ensuring that the pipeline remains portable and can scale to multi-vendor deployments. Any features that are extracted but show insufficient variation, appear too infrequently, or otherwise

lack meaningful insight for classification are subsequently discarded to reduce noise in the final dataset.

Handling gNB Traces

In contrast to the UE logs, the traces collected from the gNB are already in a more structured format, containing primarily numeric value vectors corresponding to specific timestamps. These vectors include rich information on throughput and packet transmission metrics. Because this data is already quantitative and structured, it does not require the extensive text-parsing and regular expression matching needed for the UE logs. The primary task for the gNB data is therefore selecting a subset of the most relevant features for jamming detection.

Summary of Extracted Features

Following the parsing and selection process, a final set of features is established for both the UE and gNB. These features, summarized in Table 2.2, form the basis for the subsequent interpolation and aggregation steps in the pipeline.

2.4.3 Handling Asynchronous and Missing Data via Interpolation

A significant technical challenge in fusing data from multiple, disparate log sources is handling temporal inconsistencies. Different log types are generated at unique and often irregular rates, leading to uneven timestamps and long gaps for certain metrics. Furthermore, logs can contain invalid entries due to transient hardware malfunctions that are entirely unrelated to jamming interference. These issues result in a sparse, asynchronous dataset that is unsuitable for most machine learning models, which typically require a fixed-size feature vector where all values correspond to a single, uniform point in time.

To address this, the pipeline must first create a continuous, synchronous time-series from the raw, irregular data points. This is achieved by resampling the data to a uniform interval. However, one cannot resample sparse data without first filling in the missing values. Therefore, interpolation is a critical prerequisite. This section analyzes several interpolation techniques to determine the most effective method for creating a high-fidelity, continuous signal from the discrete log entries.

Analysis of Interpolation Techniques

Several methods were considered for filling the gaps between known data points. Each technique makes different assumptions about the underlying behavior of the data:

- **Forward Filling:** This is the most straightforward method, where the last valid observation is propagated forward in time until a new data point is encountered. It operates on the logical assumption that the state of a feature persists until a new value is explicitly logged, making it well-suited for categorical or state-based data (e.g., network mode).

Feature Name	Source	Count	Range	Mean	Variance
RSSI	UE	611	(67, 89)	80.6	21.141
BER	UE	611	(64, 77)	71.767	6.913
ECSQ_2	UE	611	(-354, -306)	-327.106	72.449
ECSQ_5	UE	611	(-39, 4)	-12.315	84.105
ECSQ_6	UE	611	(-370, -318)	-339.422	111.648
ECSQ_8	UE	611	(-5, 41)	22.579	60.355
ERFTX_9	UE	112	(419, 97818)	6.48e3	1.247e8
ETHERMAL_3	UE	67	(-58, 13085)	1.052e4	2.731e7
ETHERMAL_5	UE	67	(9375, 14134)	1.278e4	4.662e6
RSRP	UE	597	(-93, -80)	-85.242	6.972
RSRQ	UE	597	(-10, -3)	-4.19	2.567
SINR	UE	597	(-2, 10)	5.273	4.047
mSleepTimeMs	UE	112	(0, 336)	3	1008
mIdleTimeMs	UE	112	(419, 97818)	6480.294	1.247e8
CQI	gNB	656	(3, 11)	7.467	1.534
PUSCH	gNB	567	(0.6, 20.1)	11.836	18.539
PUCCH	gNB	656	(0, 27.3)	6.86	22.435
DL_BRATE	gNB	656	(0, 6.5e6)	2.72e5	5.43e11
DL_OK	gNB	656	(0, 797)	33.24	8104.778
DL_NOK	gNB	656	(0, 13)	0.579	1.709
DL_PDR	gNB	656	(0, 0.32)	0.021	0.002
UL_BRATE	gNB	656	(0, 2.1e6)	8.864e5	3.384e11
UL_OK	gNB	656	(0, 228)	96.004	3963.632
UL_NOK	gNB	656	(0, 61)	5.396	37.812
UL_PDR	gNB	656	(0, 1)	0.045	0.006

Table 2.2: Feature set extracted from UE and gNB logs in a 10-minute session.

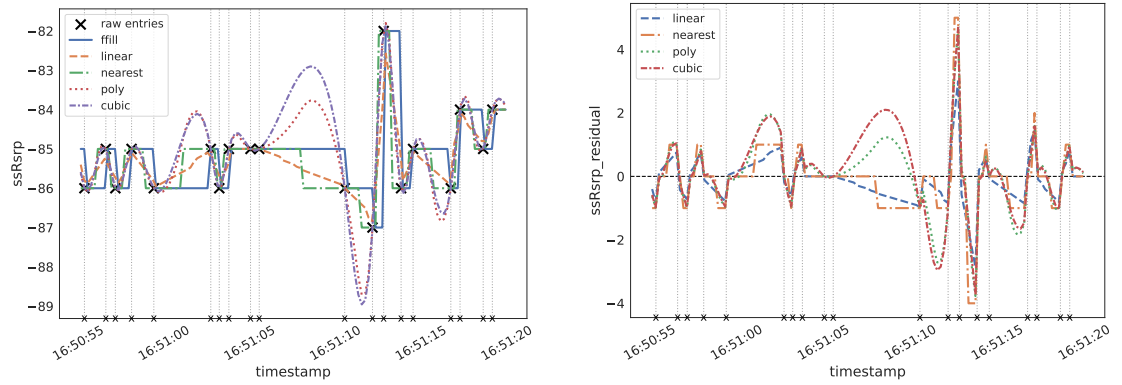
- **Nearest-Neighbor Interpolation:** This method assigns the value of the closest known data point in time to any missing point. It is computationally simple and results in a piecewise-constant function, where values change abruptly at the midpoint between two known observations.
- **Linear Interpolation:** This technique assumes a linear trend between two consecutive data points, connecting them with a straight line. Any missing value between these two points is estimated as a point on that line, calculated as a weighted average of the two known points.
- **Polynomial (Lagrange) Interpolation:** This method fits a higher-degree polynomial to a set of neighboring data points to estimate missing values. By considering multiple points, it can capture more complex curves in the data, resulting in a smoother and often more physically realistic continuous signal compared to linear methods.

- **Cubic Spline Interpolation:** This is a more advanced technique that fits a series of unique cubic (third-degree) polynomials between each pair of data points. These polynomials are constrained to join smoothly, ensuring that the resulting curve and its first and second derivatives are continuous. This produces a highly smooth and visually natural curve.

Comparative Evaluation

To select the optimal method for quantitative metrics, a visual and quantitative analysis was performed. Figure 2.5a provides a visual comparison, showing a 30-second snapshot of the RSRP feature. The plot clearly illustrates the distinct characteristics of each method: the step-like output of forward filling and nearest-neighbor, the straight-line connections of linear interpolation, and the smoother curves of the polynomial and cubic methods.

To further quantify these differences, a residual analysis was conducted, using the simple forward-filling method as a baseline reference. The residual plot in Figure 2.5b visualizes the deviation of each advanced method from this baseline, highlighting the magnitude and pattern of their differences.



(a) Time-series plot of a single feature over 30 seconds, showing the output of different methods.

(b) Residual plot comparing advanced methods against the forward-filling baseline.

Figure 2.5: Comparative analysis of interpolation techniques.

Finally, to provide an objective assessment, several standard error metrics were computed, again using forward filling as the reference. These metrics, summarized in 2.3, are defined as follows:

- **Mean Absolute Error (MAE):** The average of the absolute differences between the interpolated values and the reference values. It provides an easily interpretable measure of the average error magnitude and is less sensitive to large, isolated deviations.
- **Mean Absolute Percentage Error (MAPE):** The average absolute error expressed as a percentage of the reference values. This metric is useful for understanding the relative error but can be misleading if reference values are close to zero.

- **Root Mean Squared Error (RMSE):** The square root of the average of the squared differences. By squaring the errors, this metric gives disproportionately more weight to large errors, making it particularly sensitive to outliers.
- **R^2 :** Also known as the coefficient of determination, this metric measures the proportion of the variance in the reference data that is predictable from the interpolated data. A value closer to 1 indicates a better fit, signifying that the interpolation method closely follows the trends in the reference data.
- **Percentage Increase in Variance:** This metric calculates the percentage change in the statistical variance of the interpolated time-series compared to the variance of the reference series. It indicates how much "volatility" or "noise" an interpolation method introduces into the data.

Method	MAE	MAPE	RMSE	R^2	Var. Inc. ¹
linear	0.375	0.44%	0.635	0.936	-4.29%
nearest	0.450	0.53%	0.963	0.852	2.56%
poly	0.576	0.68%	0.962	0.853	5.47%
cubic	0.608	0.72%	0.992	0.843	6.66%

Table 2.3: Quantitative comparison of interpolation methods using forward filling as a reference.

Based on this comprehensive analysis, the SHIELD pipeline employs a two-pronged approach. For categorical features, forward filling is used due to its logical consistency with state-based data. For all quantitative metrics, polynomial (Lagrange) interpolation is selected as the optimal method. It provides a superior balance of smoothness and physical realism without the higher computational complexity of cubic splines, ensuring the resilience of subsequent ML models to data randomness and sparsity.

2.4.4 Temporal Feature Extraction via Sliding-Window Aggregation

While the instantaneous value of a metric can be informative, the temporal dynamics of that metric often contain a much richer signature of a jamming attack. A sophisticated adversary, such as the power-modulated jammer used in this work, deliberately creates interference patterns that are not easily detectable at a single point in time but become apparent when observed over a sequence of measurements. Although sequential models like Recurrent Neural Networks (RNN) are powerful for capturing such patterns [111], they often come with high computational costs that make them unsuitable for on-device deployment where resources are constrained.

To balance analytical depth with computational efficiency, we employ a sliding-window aggregation stack. This mechanism is a powerful form of feature engineering that transforms

¹Percentage Increase in Variance

the continuous time-series data into a format that captures temporal patterns without the high overhead of sequential modeling. A window of a configurable size is slid over the synchronized time-series data. Within each window, a comprehensive suite of statistical metrics is computed for each feature. This process transforms a sequence of instantaneous values into a single, fixed-size feature vector that compactly represents the data's statistical behavior over that time period. This approach filters out transient, high-frequency noise while preserving the persistent patterns indicative of an attack.

Statistical Metrics

The choice of statistical metrics is deliberate, with each one selected to capture a different aspect of the data's distribution within the window. Together, they create a high-dimensional feature space that is highly sensitive to jamming-induced anomalies.

- **Measures of Central Tendency:** These metrics (mean, median) establish the typical or central value of a feature within the window. A shift in the mean or median can indicate a sustained change in signal quality or system performance.
- **Measures of Dispersion:** These are critical for detecting the variable-power nature of the jammer. The standard deviation quantifies the volatility of a signal, while the amplitude (the difference between the maximum and minimum value) captures the range of fluctuation. A power-modulated jammer will cause these values to increase significantly, even if the mean remains relatively stable.
- **Measures of Distribution Shape:** These higher-order statistics provide deeper insights into the nature of the signal's distribution. Skewness measures the asymmetry of the distribution, which can reveal a persistent downward or upward pressure on a metric. Kurtosis measures the "tailedness" of the distribution, indicating the presence of extreme outliers. A sudden burst of interference may not significantly alter the mean but could appear as a high kurtosis value.

This aggregation is the final and most critical step in feature engineering. It creates higher-order features that are far more sensitive to the sophisticated threat model. For instance, an instantaneous RSRP value might appear normal, but a pattern of rapid fluctuation caused by the jammer will manifest as a high standard deviation within a window. This transformation is the core technical reason for the pipeline's superior detection performance.

The Balance Between Window Size and Online Detection

The size of the sliding window is a critical hyperparameter that directly influences the trade-off between detection accuracy and latency, a key consideration for online, real-time detection systems.

A **larger window** captures extended temporal trends and provides more contextual depth. This can make the system more robust to short-term, benign fluctuations and better at identifying slow-onset, subtle attacks. However, the drawback is increased latency. The system must wait for the entire window to be filled before making a prediction, which results in slower detection times.

A **smaller window** provides faster reactions and higher sensitivity to abrupt changes in the network environment. This is advantageous for detecting sudden, high-impact jamming events. However, a smaller window has less contextual information, making it potentially more susceptible to false positives triggered by transient, non-malicious network noise.

Therefore, the optimal window size must be carefully tuned to meet the specific objectives of the deployment scenario. For critical infrastructure where rapid response is paramount, a smaller window might be preferred, whereas for general performance monitoring, a larger window that prioritizes accuracy over speed might be more suitable. This tunability is a key feature of the SHIELD framework, allowing it to be adapted to diverse operational requirements.

2.5 A Rigorous Methodology for Evaluating Model Robustness

A frequent and critical shortcoming in anomaly detection research is the reliance on evaluation methodologies that can produce misleadingly optimistic results. The standard practice of applying a simple train-test split to a single, homogeneously collected dataset often leads to overly optimistic performance estimates. This is because the feature distributions in the training and testing subsets are highly similar, meaning the model is evaluated on data that closely resembles what it was trained on. Such an approach fails to capture the inherent variability of real-world production environments and, more critically, risks a subtle but significant form of data leakage, where knowledge of the test set's statistical properties inadvertently influences model design and selection. This yields models that may perform exceptionally well in the lab but are brittle and fail upon deployment.

To address this fundamental gap and provide a more rigorous assessment of model generalization, the SHIELD framework employs a two-data-collection strategy. This methodology is designed to explicitly prevent data leakage and prioritize the measurement of model resilience against new, unseen operating conditions.

1. **Primary Dataset:** The first dataset is collected under a consistent and controlled set of experimental conditions. This primary dataset is then strictly partitioned into a training set, used for fitting the machine learning models, and a validation set, used for systematic hyperparameter tuning.
2. **Robustness Test Set:** A second, entirely separate dataset is meticulously collected under demonstrably different network conditions. These variations are introduced to simulate the natural drift and change that occur in real-world environments and include factors

such as varying the UE-GNB distance, introducing new physical obstacles into the signal path, and operating under different ambient temperatures. This set is held out and used only for the final performance evaluation, ensuring a true, unbiased assessment of the model's ability to generalize.

To empirically validate the claim that the robustness set is truly different, Kernel Density Estimation (KDE) plots were generated to visualize and compare the feature distributions of the primary dataset against those of the robustness test set for both the UE and gNB. This analysis provides proof of a significant distributional shift, confirming that the robustness set represents a genuinely unseen and more challenging evaluation scenario.

2.5.1 Analysis of UE Feature Distributions

The KDE plots for the top four features from the UE dataset are shown in Figure 2.6.

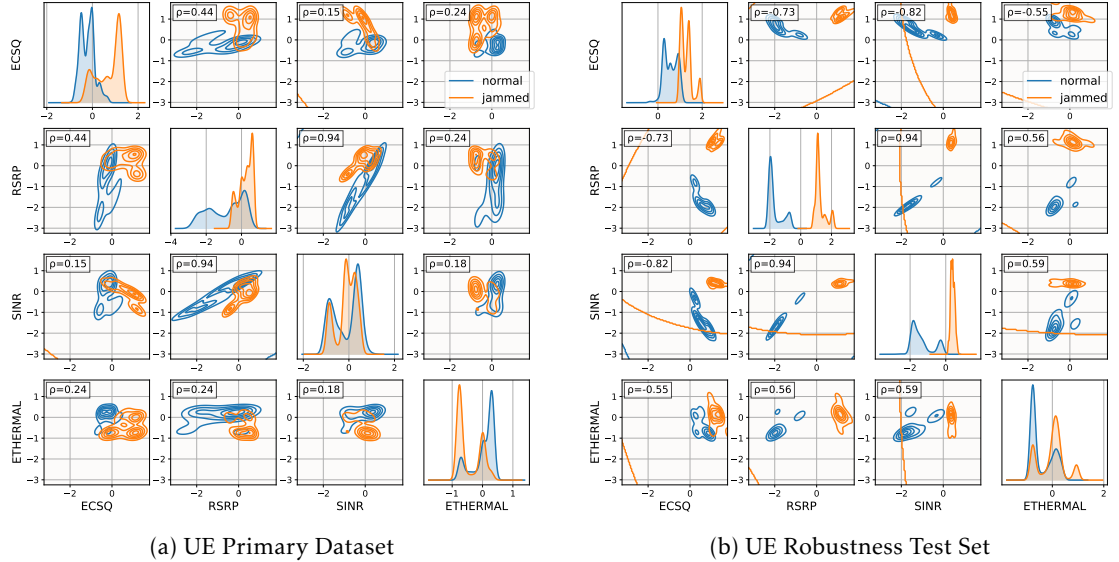


Figure 2.6: KDE analysis comparing UE feature distributions.

- **Primary Dataset 2.6a:** In the initial collection environment, there is a clear and consistent separation between the 'normal' (blue) and 'jammed' (orange) states. For instance, the SINR and RSRP distributions show distinct peaks for each class, and their strong positive correlation ($\rho = 0.94$) is maintained in both states. This indicates that within this specific environment, the jammer's impact is highly predictable.
- **Robustness Test Set 2.6b:** The distributions in the test set are markedly different. The overall shapes of the density plots for features like ETHERMAL and SINR have changed.

More importantly, the correlations between features have shifted dramatically. The relationship between ETHERMAL and SINR, which was weakly positive in the primary set ($\rho = 0.15$), is now strongly negative ($\rho = -0.82$) in the jammed state. This reversal signifies a fundamental change in the system's behavior under the new conditions.

2.5.2 Analysis of gNB Feature Distributions

A similar analysis was conducted for the top four features from the gNB dataset, shown in Figure 2.7.

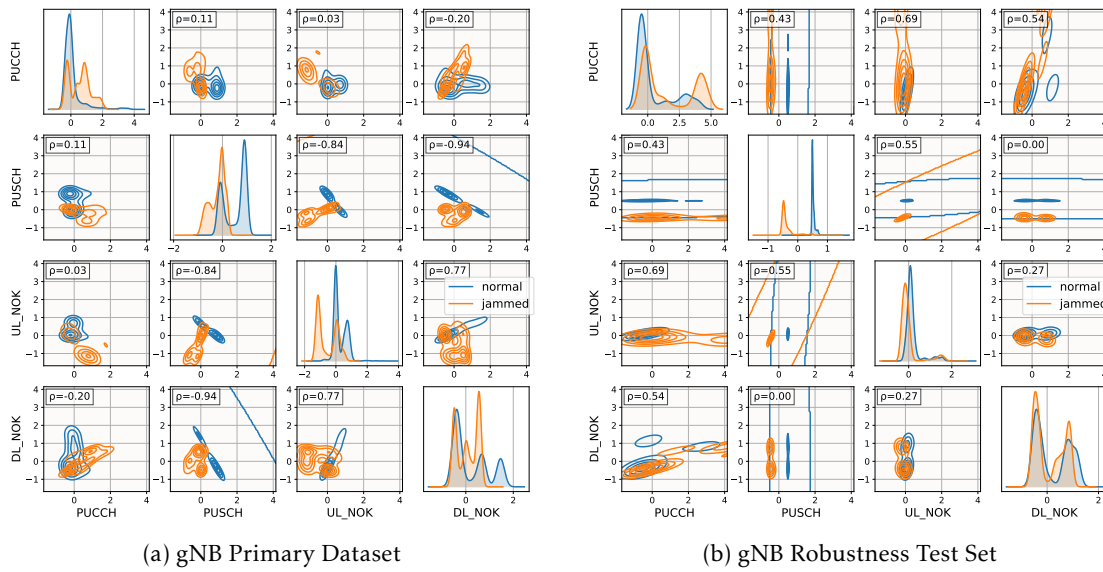


Figure 2.7: KDE analysis comparing gNB feature distributions.

- Primary Dataset 2.7a: In the primary dataset, the impact of jamming is again clear. The jammed state is characterized by strong negative correlations between key uplink metrics, such as PUSCH vs. DL_NOK ($\rho = -0.94$) and PUSCH vs. UL_NOK ($\rho = -0.84$). This suggests a predictable pattern of performance degradation.
- Robustness Test Set 2.7b: In the robustness set, these patterns are substantially altered. The correlation between DL_NOK and UL_NOK, for example, is now strongly positive ($\rho = 0.69$) in the jammed state, a stark contrast to the primary dataset. The distributions for PUSCH and PUCCH are also visibly different in shape and position.

By forcing the models to perform well on a dataset with such different statistical properties, this methodology ensures that the evaluation prioritizes true robustness over simple pattern memorization, providing a much more reliable estimate of real-world performance.

2.6 Performance Evaluation

The empirical evaluation of the SHIELD framework was conducted in two phases. First, several existing jamming detection methods from the literature were benchmarked to establish a performance baseline and demonstrate their limitations against the sophisticated threat model. Second, the full SHIELD pipeline was evaluated using a suite of lightweight ML classifiers to demonstrate its superior accuracy and robustness.

2.6.1 Benchmarking Existing Methods

To illustrate the challenges posed by the power-modulated jamming scenario, a comparative analysis was performed on several detection methods from prior literature. These methods were selected because they employ common lightweight classifiers and rely on foundational input features like PDR, Bad Packet Ratio (BPR), and basic signal strength metrics (RSSI, SNR) without the benefit of SHIELD’s preprocessing pipeline.

The results, summarized in Table 2.4, reveal a stark performance degradation when these models are moved from their original, often simulated, evaluation contexts to the more realistic and challenging environment of our testbed.

Method	Input Features	Reported Accuracy	Validation Accuracy	Test Accuracy
SVM	PDR, Packet Counts	95% [238], 98.55% [236]	73.48%	36.49%
RnF	PDR, BPR, RSSI	81% [71], 96.6% [21], 98.61% [236]	66.67%	53.65%
Hybrid-Tree	PDR, BPR, SNR, RSSI	99.5% [100]	76.51%	70.91%

Table 2.4: Comparative analysis of state-of-the-art jamming detection methods. Reported accuracies in the original studies vs. performance on our validation split and our separate robustness test set.

- Abhishek et al. [238] proposed a low-power SVM classifier operating on cumulative PDR and counts of successful or erroneous packets (Packet Counts) over 20-second intervals. Their ns-3 simulation on urban mobility scenario reported over 95% detection accuracy, even when jamming affected only 20%–40% of transmission time. When we evaluated this approach on our dataset, the accuracy dropped to 73.48%, with a notable imbalance between precision (100%) and recall (48.53%). This indicates a strong bias toward predicting attacks and overlooking normal traffic. When evaluated on our test set, the accuracy declined drastically to 36%, suggesting the SVM was unable to generalize beyond the conditions it saw in training.

- Multiple works in the literature [71, 21, 236] adopt Random Forest (RNF) classifiers using metrics such as PDR, BPR, RSSI, and SNR. While some papers briefly mention extended features, they often rely on instantaneous values rather than time-cumulative or aggregated metrics. Reported accuracies ranged from 81% to nearly 98% in simulations using ns-3 [71] or MATLAB [236]. However, our experiments under realistic conditions resulted in 66.67% validation accuracy, which dropped to 53.65% on the test scenario. This shortfall highlights the challenge of detecting subtle, power-modulated attacks that differ from the jamming profiles in simulation.
- Hone et al. [100] designed a hybrid tree-based design that combines decision trees with isolation trees to better detect unseen anomalies. This method employed fine-grained isolation trees at the leaves of the main decision tree to improve the classification of false-negative entries. They used data from C-band modems and jamming source from an Agilent E4438C signal generator, achieving near-perfect 99.5% detection in controlled tests. They further tested robustness using the open WSN-DS [15] dataset and reached 83.9% accuracy. However, their accuracy dropped to 76.51% on our validation split and 70.91% on the test scenario. Although this approach was more robust than other classifiers, it still fell substantially short of its originally reported performance.

These results starkly quantify the performance gap between controlled or simulated environments and more complex, realistic scenarios. They validate the central premise of this work: that existing methods, despite high reported accuracies, are often brittle and ill-equipped to handle sophisticated, power-modulated threats and real-world environmental variability.

2.6.2 Performance of the SHIELD Pipeline

Following the benchmarking, four common and computationally efficient classifiers—SVM, K-Nearest Neighbors (KNN), Gradient Boosting Decision Tree (GBDT), and RNF—were trained and evaluated using the feature set generated by the full SHIELD pipeline. The models were optimized using grid search with 10-fold cross-validation, with the configurations detailed in Table 2.5 reaching the optimal F1-score.

Performance on the UE Dataset

The performance of the four classifiers on the UE-derived feature set is summarized in Figure 2.8.

As shown in subplot 2.8a, all four models converged successfully on the validation split of the primary dataset, achieving high F1-scores of approximately 90-92%. The SVM model exhibited particularly strong recall (around 95%), indicating its effectiveness at minimizing false negatives in this controlled environment.

However, the true test of resilience is shown in subplot 2.8b, which details performance on the unseen robustness test set. Here, a clear performance divergence emerges. The SVM

Model	Hyperparameter	Values	F1-score (UE)	F1-score (gNB)
SVM	C	0.01 ⁺⁺ , 0.1, 10, 100	0.9147	0.9479
	kernel	linear ⁺⁺ , rbf, poly, sigmoid		
	gamma	scale, auto, 0.01, 0.1, 0.5, 1 ⁺⁺		
	degree	2, 3, 4, 5 ⁺⁺		
	tol	1e-5, 1e-4 ⁺⁺ , 1e-3		
KNN	n_neighbors	3, 5 ⁺ , 7, 9 ⁺ , 11, 15, 20	0.8475	0.9291
	weights	uniform ⁺ , distance ⁺		
	algorithm	auto ⁺ , ball_tree ⁺ , kd_tree, brute		
	p	1 ⁺⁺ , 1.5, 2		
GBDT	n_estimators	50 ⁺ , 100, 200, 300 ⁺ , 500	0.9229	0.9326
	learning_rate	1e-3, 0.01, 0.05, 0.1 ⁺ , 0.2 ⁺ , 0.3		
	min_samples_split	2 ⁺ , 5, 10 ⁺		
	min_samples_leaf	1, 2 ⁺⁺ , 5, 10		
	max_depth	3 ⁺ , 5 ⁺ , 10, 20		
	max_features	sqrt ⁺ , log2 ⁺ , None		
	subsample	0.5 ⁺⁺ , 0.7, 1.0		
RNF	n_estimators	50, 100, 200 ⁺ , 300, 500 ⁺	0.9107	0.9291
	min_samples_split	2 ⁺ , 5 ⁺ , 10		
	min_samples_leaf	1, 2 ⁺ , 5 ⁺ , 10		
	max_depth	10, 20 ⁺ , 50, None ⁺		
	max_features	sqrt ⁺ , log2 ⁺ , None		
	bootstrap	True ⁺⁺ , False		

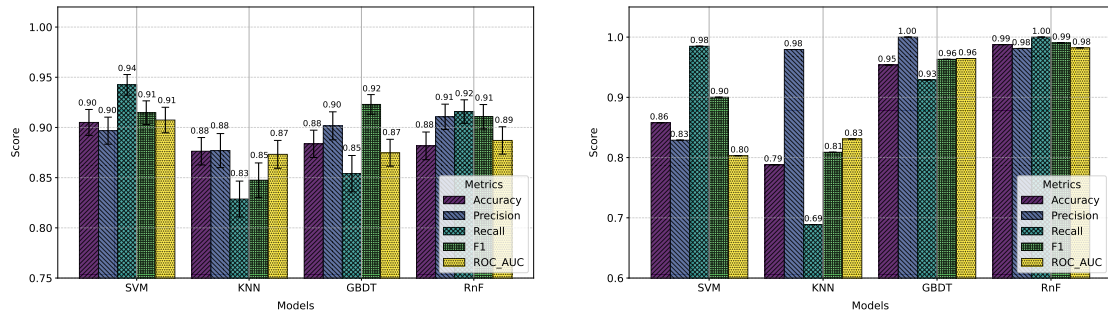
Table 2.5: Hyperparameter tuning search space and optimal configurations for each ML model. The optimal settings, determined by maximizing the F1-score during 10-fold cross-validation on the training split of the primary UE (*) and gNB (+) datasets, are presented.

and KNN models, which are more sensitive to shifts in feature distributions, experienced a notable degradation in accuracy, precision, and F1-score. In stark contrast, the tree-based ensemble models proved to be exceptionally resilient. Both GBDT and RNF maintained a high detection accuracy of approximately 94%, with F1-scores remaining robust. This demonstrates that the feature set engineered by the SHIELD pipeline, when paired with tree-based models, can effectively generalize to new and challenging conditions on the user-end device.

Performance on the gNB Dataset

A parallel evaluation was conducted using the feature set derived from gNB traces, with the results presented in Figure 2.9.

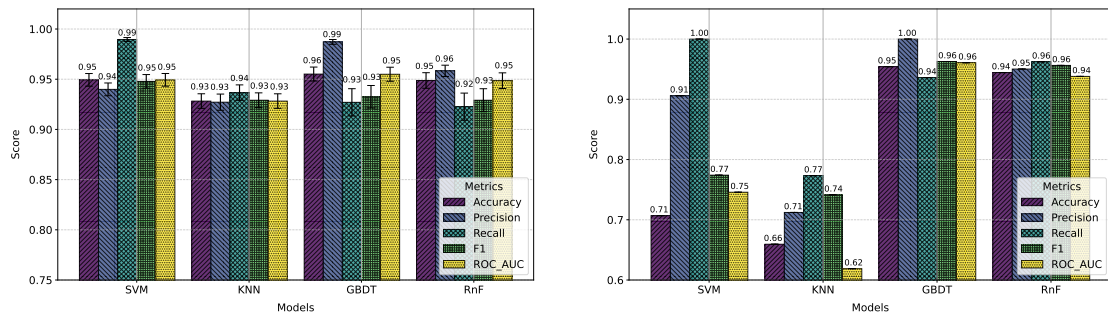
On the validation split, shown in subplot 2.9a, the models trained on gNB data achieved even higher performance than their UE counterparts, with F1-scores consistently around 93-95%.



(a) Performance on the validation split of the primary UE dataset.

(b) Performance on the separate robustness test set.

Figure 2.8: Model performance metrics on the UE dataset.



(a) Performance on the validation split of the primary gNB dataset.

(b) Performance on the separate robustness test set.

Figure 2.9: Model performance metrics on the gNB dataset.

This suggests that, under stable conditions, the gNB logs provide a slightly cleaner and more discriminative signal for jamming detection.

The evaluation on the gNB robustness test set, shown in subplot 2.9b, reinforces the findings from the UE analysis. The performance of the SVM and KNN models saw a significant decline, with the SVM's F1-score dropping to around 75%. Once again, the GBDT and RnF models demonstrated superior robustness. They successfully deal with the distributional shift, maintaining high accuracy and F1-scores of approximately 94-96%. This outstanding resilience confirms that the SHIELD pipeline is effective for infrastructure-side detection, providing a reliable mechanism for identifying jamming attacks based on the gNB's native operational logs.

2.7 Chapter Summary

This chapter has systematically addressed the critical research problem of creating a scalable, cost-effective, and robust jamming detection solution for 5G networks. It confronted the limitations of traditional hardware-dependent methods and the brittleness of prior software-based approaches

that were validated under overly simplistic conditions.

As a solution, this chapter introduced and detailed the SHIELD framework, an end-to-end system for the development and rigorous evaluation of on-device, log-based jamming detection models. The core contributions of this framework were shown to be threefold:

1. The use of a sophisticated power-modulated jammer as a realistic and challenging threat model
2. A holistic data pipeline that synchronizes and processes native logs from both the UE and gNB, eliminating the need for external hardware
3. A novel two-dataset evaluation methodology that explicitly prioritizes and measures model resilience to real-world environmental changes

The empirical evaluation provided compelling evidence of the framework's efficacy. It first demonstrated that existing detection methods from the literature fail catastrophically when confronted with the sophisticated jammer and variable network conditions of the SHIELD testbed. In stark contrast, it was shown that the SHIELD pipeline enables even lightweight ML classifiers—most notably tree-based ensembles like Random Forest and Gradient Boosting Decision Tree—to achieve and maintain a high detection accuracy of approximately 94%, even on an unseen robustness test set characterized by a significant distributional shift.

In conclusion, this chapter successfully establishes the foundational feasibility and effectiveness of on-device, log-based jamming detection. It provides the core methodological building block for the remainder of this thesis, demonstrating a practical, robust, and scalable pathway to enhancing the security of large-scale 5G and future wireless networks.

Enhancing Detector Robustness with Generative Data Augmentation

The preceding chapter established a methodology for on-device anomaly detection, demonstrating its efficacy under controlled conditions. A core finding, however, was that the performance of this detection model, like any supervised learning system, is fundamentally contingent upon the quality, diversity, and representativeness of the data used for its training. The inherent dynamism and non-stationarity of wireless environments present a formidable challenge to this dependency. The performance degradation observed when the detector encountered network conditions unseen during training underscores a critical bottleneck: the model's inability to generalize.

This chapter pivots from a model-centric to a data-centric approach to address this fundamental problem. It posits that the key to building truly robust wireless security systems lies not in further architectural refinement of the detector itself, but in solving the antecedent challenge of data scarcity and distribution shift. To this end, this chapter introduces and comprehensively validates a novel framework for generative data augmentation, designed to synthesize high-fidelity, diverse wireless time-series data. The central hypothesis is that by training a detector exclusively on a rich synthetic dataset, it is possible to imbue the model with a degree of robustness and generalization capability that is unattainable when relying solely on limited, empirically collected real-world data.

3.1 The Challenge of Data Scarcity and Model Generalization

The development of robust ML-based security solutions for wireless networks is confronted by a series of profound, data-related challenges [115]. These challenges are not merely implementation details but represent fundamental barriers that can undermine the practical viability of even

the most sophisticated detection models. This section provides a comprehensive motivation for the necessity of advanced data augmentation by deconstructing these barriers, from the logistical complexities of data acquisition to the statistical inevitability of model failure in non-stationary environments and the theoretical limitations of conventional augmentation techniques.

3.1.1 The Data Acquisition Bottleneck

The efficacy of any supervised ML model is predicated on the availability of a large-scale, high-quality, and accurately labeled training dataset. In the context of wireless security, the process of acquiring such a dataset is not a straightforward engineering task but a formidable challenge encumbered by a reinforcing system of physical, logistical, economic, and ethical constraints. These barriers collectively form a significant bottleneck that stifles the development and deployment of robust ML-based security solutions.

The Physical Barrier: The Unruly Nature of the Wireless Medium

The physical nature of the wireless medium creates a fundamental problem of data integrity and reproducibility for ML applications, making comprehensive data collection nearly intractable. A wireless signal's characteristics are subject to a vast and continuously changing array of environmental variables [182, 137]. These include static physical obstructions like buildings, dynamic propagation effects such as multipath fading, interference from other wireless technologies, and atmospheric conditions like temperature and humidity [262].

This volatility means that even identical, benign transmissions can result in vastly different input vectors for an ML model if sent at different times or from slightly different locations [155, 153]. Research has even shown that minor changes in an indoor environment, such as displacing furniture by a few centimeters, can produce statistically significant changes in Channel State Information (CSI) data [234]. Consequently, this introduces enormous intra-class variance, making it exceedingly difficult for a model to learn a stable and generalizable decision boundary for normal behavior. Essentially, the definition of normal is in a constant state of flux, driven by the laws of physics rather than a change in semantics.

Furthermore, the wireless channel is not a noiseless digital medium but a complex and chaotic physical environment. Unlike wired networks where the communication channel is relatively stable and protected, wireless signals propagate through open space, subjecting them to a host of unpredictable transformations that fundamentally alter the data before it ever reaches a receiver's ML model [146]. Signals undergo reflection, attenuation, and phase shifts as they interact with the environment, a phenomenon known as multipath fading [249]. Consequently, the input vector observed by an ML model is a complex convolution of the originally transmitted signal and the time-varying channel impulse response. This physical transformation is a critical yet often overlooked aspect of data acquisition. The "ground truth" of a transmitted signal is never what the model actually sees.

The Economic Barrier: The Prohibitive Cost of Data

Beyond the physical and device-level constraints, the data acquisition process is fundamentally governed by economic realities. Building large-scale, high-quality datasets for wireless security is an expensive and resource-intensive endeavor [250, 89]. The process requires significant capital investment in specialized hardware and software for data collection, storage, and analysis. High-quality wireless data collection often necessitates the extensive deployment of specialized equipment, a process that is both costly and time-consuming [127, 185]. This financial burden is compounded by the substantial human capital required. The process of labeling network traffic, a critical step for supervised learning, is particularly costly as it demands the time of highly specialized security experts [84, 198].

Furthermore, the operational costs are heavily influenced by the energy consumed during data collection and transmission. In environments like WSN, energy is a critical and finite resource, with sensor nodes often powered by small batteries that are difficult or impossible to replace [61, 18]. The act of collecting and transmitting large volumes of data is one of the most energy-intensive operations for these devices, directly impacting their operational lifespan and increasing long-term maintenance costs [200]. This creates a direct economic trade-off: the pursuit of more data for better security models leads to higher energy expenditure and potentially shorter network lifetimes, which in turn necessitates more frequent and costly hardware replacement [10]. The economic calculus of data acquisition, therefore, involves balancing the need for comprehensive data against the tangible costs of hardware, expert personnel, and sustained energy consumption.

Table 3.1 further highlights the economic considerations involved in hardware selection for data acquisition in experimental environments.

The Compliance Barrier: Ethical and Legal Constraints

The collection and analysis of network traffic data are activities fraught with significant ethical and legal complexities. Global data privacy regulations, such as the GDPR in Europe, impose stringent requirements on how organizations handle personal data [240]. These frameworks mandate principles of transparency, user consent, data minimization, and robust security, creating substantial operational and financial overhead for any entity seeking to collect network data for security analysis [76, 104].

Network traffic is a rich source of potentially sensitive information. Packet data can contain Personally Identifiable Information (PII), confidential business communications, and user behavior patterns [160]. Even high-level flow data, which summarizes connections without inspecting packet content, can reveal sensitive associations and activities. Therefore, any form of network traffic analysis must be conducted within a strict legal and ethical framework to avoid violating individual privacy rights [168]. Organizations must be transparent with users about what data is being collected and for what purpose, and in many cases, must obtain explicit and informed

Device	Price (USD)	Power Source	Power Consumption	Notes/Context
RTL-SDR Dongles	~\$30-50	USB 2.0	~110-300mA (0.55-1.5W)	Consumption varies by tuner chip; R820T tuners draw more power than E4000 tuners.
LimeSDR (Mini)	~\$139 / ~\$440 (v2.0)	USB 3.0	~200mA (1W)	Lower consumption compared to the full LimeSDR board.
PlutoSDR	~\$230	USB 2.0	~411mA at 5V (~2.07W)	Consumption rises to ~2.2W in full-duplex operation.
HackRF One	~\$320	USB 2.0	Requires 500mA at 5V (2.5W)	A powered USB hub may be necessary if the host PC cannot meet the 500mA requirement.
BladeRF 2.0 micro	\$540 (xA4)	USB 3.0 / External DC	~600mA at 4.78V (~2.88W)	An external power supply is recommended for peripherals.
Ettus USRP B210	\$2,165	USB 3.0 / External DC	1.9-4.1W	Power draw is dependent on the number of RX/TX channels, sample rate, etc.
Ettus USRP E320	~\$9,260	External DC (10-14V)	Up to 30W	An embedded series device designed for more demanding applications.
Ettus USRP X310	~\$11,200	External DC	~45W	High-performance model; consumption figure includes two SBX-120 daughterboards.
Ettus USRP X410/X440	Starting at ~\$29,000	External DC (12V)	Up to 192W (16A max)	High-end, high-performance devices with substantial power requirements reflecting their advanced capabilities.

Table 3.1: Qualitative Analysis of Commercial Software-Defined Radio (SDR) Devices from an Economic Perspective.

consent. Furthermore, they are accountable for securing this data against breaches, with severe financial and reputational penalties for failure.

This creates a fundamental conflict between the requirements for effective security monitoring and the mandate for privacy protection. To build a highly accurate ML model for intrusion detection, analysts often need access to granular data, including features that may be considered PII. Standard privacy-preserving techniques, such as data anonymization or aggregation, can resolve the compliance issue but often at the cost of data utility. By removing or obscuring

critical features like source IP addresses or specific port numbers, these techniques can blind the ML model to the subtle indicators of a sophisticated attack. This results in a direct and challenging trade-off: maximizing data fidelity for security can increase privacy and compliance risks, while maximizing privacy can degrade the model's ability to detect threats. This impasse is a powerful motivator for the use of high-fidelity synthetic data, which can be engineered to mirror the statistical properties of real network traffic without containing any actual PII, thereby satisfying both security and privacy requirements [65, 66, 144].

3.1.2 Distribution Shift and the Failure of Generalization

A foundational assumption in many ML applications is that the data used for training and the data encountered during deployment are independent and identically distributed. However, the non-stationary nature of wireless environments systematically violates this assumption, leading to a phenomenon known as distribution shift [130, 261]. This shift is the statistical manifestation of the physical dynamism described previously and is a primary cause of generalization failure in deployed models [255].

Distribution shift can be formally characterized into two principal types:

- **Covariate Shift:** This occurs when the marginal distribution of the input features, denoted as $\mathbb{P}(X)$, changes between the training and testing domains, while the conditional probability of the label Y given the features, $\mathbb{P}(Y | X)$, remains constant. In the context of wireless security, a change in the physical environment—such as a user moving to a new location, the introduction of a new piece of furniture, or a change in weather—alters the statistical properties of the received signal features X [155]. The underlying relationship between a specific signal pattern and its label (e.g., 'anomaly') might not change, but the features themselves now occupy a different region of the input space.
- **Concept Drift:** This refers to a change in the conditional distribution $\mathbb{P}(Y | X)$, meaning the fundamental relationship between the features and the target label evolves over time. This can occur, for instance, if an adversary alters their attack methodology [192, 242]. A jamming signal that previously had a specific waveform might be replaced with a new one, meaning the same set of input features X that once corresponded to a benign state ($Y = \text{'normal'}$) now signifies an attack ($Y = \text{'anomaly'}$) [184].

The physical phenomena inherent to wireless communication are the direct causal drivers of these statistical shifts. A change in atmospheric humidity is a physical event that perturbs the data distribution $\mathbb{P}(X)$ [110, 33]. An adversary deploying a new attack vector is a strategic action that perturbs the conditional concept distribution $\mathbb{P}(Y | X)$. Consequently, the failure of a ML model to generalize is not an abstract or unpredictable problem. It is a statistically guaranteed outcome of the physics of the wireless medium. This continuous induction of distribution shifts ensures that any model trained on a static dataset is destined to become suboptimal. This

provides a rigorous, first-principles justification for developing solutions that are inherently designed to be robust to such shifts, rather than assuming a stable operational environment.

The empirical evidence of this phenomenon was observed in the baseline experiments of this thesis, where the detector's accuracy plummeted from approximately 92% on in-distribution data (Scenario A) to just 78% on out-of-distribution data (Scenario B), quantifying the severe impact of this generalization gap [257].

3.1.3 The Inadequacy of Traditional Augmentation Techniques

Given the immense challenges of data acquisition and the inevitability of class imbalance in security datasets, where malicious instances are typically far rarer than benign ones [117], data augmentation has become a standard practice in the ML workflow. The goal is to artificially expand the training dataset to improve model robustness and generalization. However, the traditional techniques commonly employed are often fundamentally mismatched with the complex, high-dimensional, and temporally structured nature of wireless security data. These methods, ranging from simple interpolation to domain-agnostic transformations, can not only fail to help but may actively harm model performance.

A Critique of Interpolation-Based Oversampling (SMOTE)

The Synthetic Minority Over-sampling Technique (SMOTE) [42] is one of the most widely used methods for addressing class imbalance. Its core mechanism is to generate new synthetic samples for the minority class by selecting an existing minority instance, identifying its k -nearest neighbors, and creating a new data point along the line segment connecting the instance to one of its neighbors [22]. This approach is intended to create more varied minority class data than simple replication, thereby reducing overfitting. Numerous studies in intrusion detection [12, 70] have shown that applying SMOTE can lead to improved classification metrics like accuracy, precision, and recall by balancing the dataset before training.

Despite its popularity, SMOTE and its variants suffer from several critical flaws that limit their effectiveness, particularly in complex feature spaces. A primary issue is that the interpolation process can generate unrealistic or noisy samples. By creating new points in-between existing ones, SMOTE can blur the decision boundary between the minority and majority classes, making it harder for the classifier to learn a clean separation. In cases where the classes are already well-separated, studies have shown that SMOTE can significantly degrade model performance by introducing these ambiguous, boundary-blurring samples [227].

Furthermore, the synthetic samples generated by SMOTE are not truly novel. They are merely simple linear combinations of existing data. This can still lead to a form of overfitting, where the model learns the specific artifacts of the SMOTE algorithm rather than the true underlying distribution of the minority class [251]. Another subtle but significant drawback is that by artificially altering the data distribution in specific regions of the feature space, SMOTE can lead

to poorly calibrated models [172]. The model's output probabilities may no longer reflect the true likelihood of a class, which is problematic for security applications where confidence scores are used for risk assessment and alert prioritization.

When applied to wireless data, these problems are magnified. A data point in wireless security represents a physical signal or a network event governed by specific protocols and physical laws. The features in this data point (e.g., SNR, packet inter-arrival time, protocol flags) are not independent variables that can be arbitrarily interpolated. The linear combination of two valid, physically-realizable data points does not guarantee a new, physically-realizable data point. SMOTE operates in the feature space without any regard for the underlying generative process of the data, meaning it can create synthetic samples that correspond to "impossible" physical signals or nonsensical network events. Training a model on such semantically invalid data forces it to learn spurious correlations, ultimately hindering its ability to generalize to real-world scenarios.

The Fallacy of Domain-Agnostic Transformations

Another category of traditional data augmentation involves applying simple transformations to existing data points. Many of these techniques have been borrowed directly from the domain of computer vision, where they have proven highly effective [116]. These include geometric transformations like rotation, cropping, flipping, and translation, as well as color space transformations and noise injection. The success of these methods in image processing relies on a key property: object identity is often invariant to these transformations. A rotated or slightly brighter image of a cat is still recognizably a cat.

However, applying this logic to wireless security data is a fundamental fallacy. Wireless data, whether it be raw I/Q samples from the physical layer or packet-level features from network flows, is typically high-dimensional time-series data [249]. In this domain, the information is not just in the magnitude of the values but is critically encoded in their temporal sequence and structure [77, 105]. Domain-agnostic transformations violate this structure and destroy the semantic meaning of the data.

Consider the analogy of a TCP handshake, which is defined by a specific, ordered sequence of packets (SYN, SYN-ACK, ACK). Applying a random permutation or shuffling augmentation, which is analogous to cropping and rearranging patches of an image, would destroy this required sequence, transforming a valid network connection into a stream of meaningless noise [105]. Similarly, applying a rotation transformation to a time-series signal is not a well-defined or meaningful operation. Adding random jittering or noise to the timing features of packets could inadvertently change the interpretation of the data from a benign flow to what appears to be a timing-based side-channel attack.

These transformations manipulate the data as a generic array of numbers without respecting its underlying physical laws [249]. As a result, they do not create new, valid examples of a given class. Instead, they create corrupted, invalid, or even mislabeled data points that actively

poison the training set and teach the model incorrect patterns. This fundamental mismatch demonstrates that effective data augmentation for wireless security cannot be a generic, borrowed process. It requires a domain-aware approach that can understand and replicate the complex temporal and structural properties of the data, a task for which simple transformations are inadequate.

The fundamental differences between these two paradigms are outlined in Table 3.2.

Attribute	Traditional Augmentation (e.g., SMOTE, Transformations)	Generative Augmentation (e.g., GAN, VAE, Diffusion)
Core Principle	Transformation or interpolation of existing data points.	Learning the underlying data distribution to synthesize novel samples.
Data Fidelity	Low to Medium. Can create physically and semantically impossible samples by violating domain constraints (e.g., temporal order, physical laws).	High. Aims to generate samples that are statistically indistinguishable from the real data distribution, preserving complex correlations.
Diversity Generation	Limited. Creates new samples that are linear combinations or simple variations of existing data, often staying close to the original data manifold.	High. Capable of generating entirely new modes and variations within the data distribution, exploring unseen but plausible data points.
Domain Awareness	Low. Generally agnostic to the data's underlying structure, such as temporal dependencies in time-series data or protocol rules in network traffic.	High. Models like GAN, VAE, and Diffusion Models can learn complex, non-linear relationships and domain-specific rules implicitly from the data.
Overfitting Risk	High. Replicating or creating simple variations of minority samples can lead to the model memorizing these specific instances and their artifacts.	Medium to Low. By generating diverse new samples, it encourages better generalization. However, a poorly trained generator can still produce artifacts.
Suitability for Wireless Data	Poor. Methods like random shuffling or rotation destroy critical temporal structure. Interpolation methods like SMOTE can create physically invalid signal representations.	Excellent. Capable of modeling the complex, high-dimensional, and time-series nature of wireless signals and network traffic flows.

Table 3.2: A Comparative Analysis of Traditional and Generative Data Augmentation Techniques.

3.1.4 The Case for Generative Models: A New Paradigm for Data Synthesis

The systemic challenges of data acquisition, the inevitability of concept drift, and the profound inadequacy of traditional augmentation techniques necessitate a fundamental shift in how datasets for wireless security are created and maintained. This shift moves away from the paradigm of simply collecting or transforming data and towards the paradigm of generating it. Generative Artificial Intelligence, particularly deep generative models, offers a powerful solution by learning the complex, underlying distributions of real-world data and using that knowledge to synthesize novel, high-fidelity, and diverse data points on demand.

Principles of Generative Data Augmentation

At its core, generative data augmentation operates on a principle fundamentally different from that of traditional methods. Instead of taking existing data points and applying simple modifications (e.g., adding noise, rotating, interpolating), generative models aim to build a comprehensive statistical model of the data itself [82]. The objective is to learn the joint probability distribution $\mathbb{P}(Y | X)$, where X represents the features and Y the class label. Once this distribution is learned, the model can be sampled to produce entirely new data points $(\hat{X}, \hat{Y})$ that are statistically consistent with the original dataset.

This approach directly addresses the core data-centric challenges. For data scarcity and class imbalance, generative models can be used to synthesize a virtually unlimited number of high-quality samples for underrepresented classes, such as rare attack types [249]. Unlike SMOTE, which creates simple linear combinations, deep generative models can capture the complex, non-linear relationships between features, resulting in synthetic data that is far more realistic and diverse [113, 48]. This helps to balance the dataset without the risk of simple duplication and overfitting that plagues traditional oversampling methods.

For privacy and compliance, generative models offer a transformative solution. They can be trained on sensitive, real-world network data and then used to generate a synthetic dataset that preserves the statistical patterns and properties of the original data but contains no actual PII [113]. This allows researchers and organizations to develop, test, and share robust ML models without exposing sensitive user information, breaking the privacy-fidelity trade-off [81].

Furthermore, generative models provide a powerful tool for proactive security. They can be used to simulate rare anomalies and novel attack scenarios that may not yet exist in collected datasets [86]. By exploring the learned latent space of threats, security analysts can generate what-if scenarios, stress-testing their defenses against hypothetical future attacks and moving from a reactive to a proactive security posture. This ability to create high-volume and labeled datasets on demand transforms the problem from one of data scarcity to one of data abundance, shifting the focus from "how do we find more data?" to "how do we build a better data generator?".

A Comparative Analysis of Generative Architectures

The field of generative modeling is dominated by three principal families of architectures: GAN, Variational Autoencoder (VAE), and Diffusion Models. Each has a unique mechanism and a distinct set of strengths and weaknesses, making them suitable for different tasks within the wireless security domain. The distinct trade-offs between these architectures are summarized in Table 3.3.

Generative Adversarial Networks A GAN architecture consists of two neural networks, a Generator and a Discriminator, locked in a competitive, zero-sum game. The Generator's role is to take a random noise vector as input and transform it into a synthetic data sample that mimics the real data. The Discriminator's role is to act as an expert, receiving both real samples from the training set and fake samples from the Generator, and learning to distinguish between the two. During training, the Generator continuously improves its ability to create more realistic samples to fool the Discriminator, while the Discriminator becomes better at detecting fakes. This adversarial process continues until a Nash equilibrium is reached, where the Generator produces samples that are so realistic the Discriminator can no longer tell them apart from real data with better than chance accuracy.

- **Strengths:** GANs are renowned for their ability to produce exceptionally high-quality, sharp, and realistic outputs, often surpassing other generative models in fidelity [40]. This makes them highly effective for generating synthetic data that must closely resemble real-world signals or network traffic. Their adversarial nature is also conceptually well-aligned with cybersecurity, where they can be used to simulate sophisticated attacks or generate adversarial examples to harden defenses [9]. Another significant advantage is their fast inference speed. Once trained, a GAN's generator can produce a new sample in a single forward pass, making it suitable for real-time or large-scale data generation tasks.
- **Weaknesses:** The primary drawback of GANs is their training instability. The delicate balance required between the Generator and Discriminator can be difficult to achieve, often leading to training divergence, or a phenomenon known as "mode collapse" [239]. Mode collapse occurs when the Generator finds a few specific outputs that are highly effective at fooling the Discriminator and begins producing only those limited variations, failing to capture the full diversity of the true data distribution.

Variational Autoencoders A VAE is a probabilistic generative model built upon the autoencoder architecture. It consists of two parts: an Encoder network that compresses high-dimensional input data into a low-dimensional, probabilistic latent space, and a Decoder network that reconstructs the original data from a point sampled from this latent space [121]. Unlike a standard autoencoder, a VAE's encoder outputs the mean and variance of a gaussian probability distribution for each input. The training objective is twofold: to minimize the reconstruction loss

Metric	GAN	VAE	Diffusion
Training Stability	Unstable. Requires careful tuning and balancing of two competing networks. Prone to issues like mode collapse.	Stable. Optimized with a single, well-defined loss function, making training straightforward and reliable.	Stable. The training objective is a well-behaved denoising score matching process, avoiding adversarial dynamics.
Sample Quality (Fidelity)	Excellent. Produces sharp, highly realistic samples that can be indistinguishable from real data.	Good to Moderate. Often produces smoother or blurrier samples due to the averaging effect of the probabilistic latent space.	State-of-the-Art. Capable of generating samples with exceptional detail, quality, and realism, often surpassing GAN.
Sample Diversity	Good, but can be compromised by mode collapse, where the generator produces only a limited variety of samples.	Excellent. The probabilistic and continuous nature of the latent space encourages the generation of diverse and varied samples.	Excellent. The stochastic, iterative generation process naturally leads to a wide diversity of high-quality outputs.
Inference Speed	Very Fast. Generation requires only a single forward pass through the trained generator network.	Fast. Generation requires a single forward pass through the trained decoder network.	Very Slow. Generation is an iterative process requiring thousands of forward passes through the denoising network.
Latent Space Control	Moderate. The latent space can be unstructured, making controlled generation difficult without specialized architectures (e.g., cGAN, InfoGAN).	Excellent. The latent space is explicitly regularized to be smooth and continuous, allowing for meaningful interpolation and manipulation.	Moderate to Good. Latent space control is an active area of research, but techniques exist to guide the generation process.
Cybersecurity Use Case	Generating high-fidelity adversarial examples; simulating realistic but known attack patterns; fast, real-time data augmentation for IDS.	Anomaly detection (by modeling normal data distribution); generating structured tabular data; controllable synthesis of threat variations.	High-fidelity, privacy-preserving generation of entire network traffic datasets; augmenting rare attack classes with maximum realism and diversity.

Table 3.3: A Comparative Overview of Key Generative Model Architectures.

(how well the decoder can recreate the input) and to minimize the Kullback-Leibler divergence (KL DIVERGENCE) between the learned latent distribution and a prior distribution (e.g., a standard normal distribution). This second objective acts as a regularizer, forcing the latent space to be smooth and well-structured.

- **Strengths:** VAEs are significantly more stable to train than GAN because they are optimized via a single, well-defined loss function rather than a competitive game [239]. Their well-structured, continuous latent space is a major advantage for tasks that require smooth interpolation between data points or controlled, interpretable data generation. This makes them particularly useful for anomaly detection, where a VAE can be trained on normal data, and anomalies can be identified as inputs that have a high reconstruction error, indicating they do not fit the learned distribution of normality [222].
- **Weaknesses:** The primary trade-off with VAE is sample quality. The probabilistic nature of the encoder and the pressure from the KL DIVERGENCE term often lead to the generation of outputs that are blurrier, smoother, or less detailed than those produced by GAN [40]. This is a result of the model averaging over the data distribution to satisfy the reconstruction and regularization objectives. While suitable for many tasks, VAE may struggle to capture the fine-grained, sharp details necessary for generating highly realistic attack signatures or complex signal data.

Diffusion Models Diffusion Models [99] are a newer class of generative models that have recently achieved state-of-the-art results in many domains, particularly image synthesis [13]. They operate based on a two-step process. The forward process involves systematically and gradually adding Gaussian noise to a real data sample over a series of time steps, until the sample is transformed into pure, unstructured noise. The reverse process then trains a neural network to reverse this process. To generate a new sample, the model starts with random noise and iteratively applies the trained denoising network over the same number of time steps, gradually refining the noise into a clean, structured data sample.

- **Strengths:** Diffusion models are capable of producing samples of exceptionally high quality and diversity, often exceeding the performance of even the best GAN [230]. Their training process is stable and well-defined, avoiding the adversarial dynamics that plague GAN. This combination of quality, diversity, and stability makes them a very promising architecture for generating highly realistic and varied synthetic data for wireless security applications.
- **Weaknesses:** The main disadvantage of diffusion models is their computationally intensive training process and slow inference speed. The iterative denoising process requires multiple forward passes through the neural network to generate a single sample [158]. This makes them computationally expensive and much slower at inference time compared to GAN or VAE, which can be a significant limitation for applications requiring rapid, on-demand data generation.

3.2 The GANSec Framework for Wireless Time-Series Augmentation

To address the data-centric challenges outlined, this work proposes GANSec, a tailored framework based on conditional GAN designed specifically for augmenting labeled wireless time-series data [257]. The framework's objective is to generate high-fidelity synthetic data that not only captures the complex temporal and statistical properties of real wireless signals but also improves the robustness and generalization of downstream anomaly detectors trained upon it. The overall workflow is depicted in Figure 3.1.

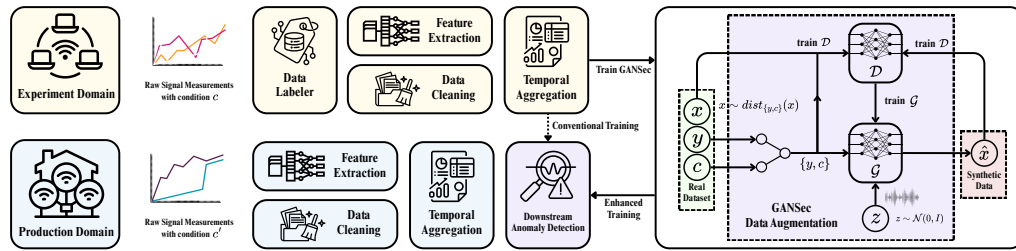


Figure 3.1: GAN-based Data Augmentation for Anomaly Detection in Wireless Security: Workflow Overview.

The selection of GAN as the foundation for the framework proposed in this thesis is a pragmatic and well-reasoned decision. The downstream task of anomaly detection demands samples with high fidelity to preserve the subtle characteristics that differentiate normal from anomalous signals, a domain where GANs excel over VAEs. While diffusion models may offer superior quality, their slow sampling speed is a practical impediment for data augmentation scenarios that may require the rapid generation of thousands or millions of samples to construct a balanced and diverse training set. Therefore, GANs represent an optimal trade-off, providing the requisite sample quality for a discriminative task with a more efficient generation process than diffusion models.

3.2.1 Conditional GAN Fundamentals

GANSec is built upon the Conditional GAN (cGAN) paradigm, which extends the original GAN architecture to allow for guided and controlled data generation [157]. In a standard GAN, the generator (denoted as $\mathcal{G}$) learns a mapping from a random noise vector z to the data space X , expressed as $\mathcal{G} : z \rightarrow x$. This process is unconditional, offering no direct control over the characteristics of the output sample.

A cGAN modifies this by providing both the generator and the discriminator (denoted as $\mathcal{D}$) with additional conditioning information, typically a class label y or a more general context vector c . The generator thus learns a mapping $\mathcal{G} : \{z, y, c\} \rightarrow x$, producing a sample

that corresponds to the given condition. The $\mathcal{D}$, in turn, learns to assess whether a sample x is authentic given the label y and the condition c , i.e., $\mathcal{D}(x | y, c)$. This mechanism enables explicit control over the generation process, making it possible to synthesize data belonging to a specific class (e.g., 'normal' or 'jammed') and under specific contextual conditions (e.g., device location, modem temperature), which is essential for building labeled datasets for supervised learning [220].

The training objective remains an adversarial minimax game, but with the inclusion of the conditioning vector. The value function $\mathbb{V}(\mathcal{D}, \mathcal{G})$ is formulated as:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{V}(\mathcal{D}, \mathcal{G}) = \mathbb{E}_{x \sim p_{\text{data}}}(x) + \mathbb{E}_{z \sim p_z}(z)$$

Here, the generator $\mathcal{G}$ tries to minimize the probability of the discriminator correctly identifying its outputs as fake, while the discriminator $\mathcal{D}$ tries to maximize this probability. Through this adversarial process, $\mathcal{G}$ learns to produce samples that are indistinguishable from real data corresponding to the specified condition $\{y, c\}$.

3.2.2 Generator and Discriminator Design

The effectiveness of a GAN in modeling a specific data type is heavily influenced by the architectural choices for its generator and discriminator networks. Recognizing that wireless data is inherently sequential, GANSec explores three distinct neural network backbones, each embodying a different set of assumptions, or inductive biases, about the underlying structure of the data. This exploration is not merely a search for the best architecture but also an implicit investigation into which features are most critical for characterizing the wireless anomalies under study.

- **Multi-Layer Perceptron (MLP):** This architecture serves as a standard, fully-connected baseline. To process the 2D time-series tensors (with dimensions of time steps and features), the input must be flattened into a 1D vector. This process inherently discards the explicit temporal ordering and structure of the data. An effective MLP-based GAN would imply that the anomalous behavior can be identified purely from the statistical distribution of feature values, without regard to their sequence.
- **Long Short-Term Memory (LSTM):** LSTMs are a specialized form of RNN designed to overcome the vanishing gradient problem, enabling them to effectively learn long-range dependencies in sequential data. An LSTM-based backbone operates under the hypothesis that the defining characteristics of a wireless anomaly are embedded in its temporal evolution and the complex, long-term correlations between signal measurements over time.
- **CNN:** While commonly associated with image processing, 1D or 2D CNNs can be powerfully applied to time-series data. In this context, the convolutional filters learn to detect

local, time-invariant patterns within the signal sequence. A successful CNN-based GAN would support the hypothesis that the anomaly is best characterized by specific, short-term signal shapes or signatures that can appear at any point within the observation window.

The eventual performance of detectors trained on data from these different GAN architectures provides empirical evidence for these competing hypotheses. As will be shown in the evaluation, the superior results achieved with the CNN backbone suggest that for the power-modulated jamming anomaly investigated, the presence of local, translation-invariant patterns is the most salient feature for both generation and subsequent detection.

3.2.3 Fidelity-focused vs. Task-oriented Generation

A central innovation within the GANSec framework is the systematic investigation of two distinct training objectives, which steer the generation process towards different goals. This design choice exposes a fundamental trade-off in generative modeling for downstream tasks, which can be conceptualized as a balance between distributional fidelity and task-specific utility.

Embedded Conditional (EC) Objective

In the EC-GAN configuration, the primary objective is to generate realistic data samples that accurately match the provided conditions $\{y, c\}$. This approach prioritizes fidelity, compelling the generator $\mathcal{G}$ to learn a model of the true conditional data distribution, $\mathbb{P}(X | Y, C)$, as accurately as possible.

As shown by the 'EC-Discriminator' path in Figure 3.2, the conditioning information is embedded and fed into the discriminator $\mathcal{D}$ alongside the data sample. The $\mathcal{D}$'s role is singular: to assess the authenticity of a given sample. It takes a data sample x and its corresponding condition vector $\{y, c\}$ as input and outputs a single probability score indicating whether the sample is real or synthetic.

The $\mathcal{G}$'s loss function is therefore driven exclusively by its ability to spoof this $\mathcal{D}$. In turn, it aims to produce data that $\mathcal{D}$ classifies as real for the specified condition. The goal is to generate samples that are statistically indistinguishable from real-world data. The training process is detailed in Algorithm 2.

Classification Oriented (CO) Objective

The CO-GAN configuration extends the discriminator $\mathcal{D}$'s role beyond simple real/fake classification, explicitly biasing the generator $\mathcal{G}$ towards task-utility.

As depicted by the 'CO-Discriminator' path in Figure 3.2, $\mathcal{D}$ simultaneously performs two tasks: it predicts whether the input data is real or synthetic, and it predicts the data's class label y based on the condition c . Consequently, the $\mathcal{G}$'s loss function becomes a composite of two terms: one that rewards the $\mathcal{G}$ for creating samples that $\mathcal{D}$ deems 'real', and a second that rewards it for creating samples that $\mathcal{D}$ also classifies with the correct target label y . The intuition

Algorithm 2 Embedded Conditional GAN Training: $\mathcal{D}$ focuses on distinguishing real samples from synthetic samples based on the provided context

Require: Feature set $x \sim \text{dist}_{\{y,c\}}(x)$ with label y and condition c

Ensure: Discriminator $\mathcal{D}$, Generator $\mathcal{G}$

- 1: Initialize number of epochs ϵ , loss function $\mathcal{L}$, loss weights ω_i
- 2: Concatenate $\{y, c\}$
- 3: **for** current_epoch $\leftarrow 1, 2, \dots, \epsilon$ **do**
- 4: Sample noise $z \sim \mathcal{N}(0, I)$
- 5: Generate synthetic data using $\mathcal{G}$: $\hat{x} \leftarrow \mathcal{G}(z, \{y, c\})$
- 6: Discriminate real or synthetic data using $\mathcal{D}$

$$\text{output_real} \leftarrow \mathcal{D}(x, \{y, c\})$$

$$\text{output_synthetic} \leftarrow \mathcal{D}(\hat{x}, \{y, c\})$$

- 7: Compute loss for $\mathcal{G}$ and $\mathcal{D}$

$$\text{loss}_{\mathcal{D}} \leftarrow \omega_0 \times \mathcal{L}(1, \text{output_real}) + \omega_1 \times \mathcal{L}(0, \text{output_synthetic})$$

$$\text{loss}_{\mathcal{G}} \leftarrow \omega_2 \times \mathcal{L}(1, \text{output_synthetic})$$

- 8: **end for**
-

is that by optimizing for class discriminability alongside realism, $\mathcal{G}$ learns to generate samples that are particularly useful for the downstream classification task, potentially enhancing the anomaly detector's performance. The training process is detailed in Algorithm 3.

This duality presents a nuanced choice. The EC approach may generate a more diverse and truly representative set of samples, while the CO approach might learn to generate caricatures of each class. These are samples that are exaggeratedly distinct and easy to classify, potentially at the expense of perfectly matching the real data distribution. The empirical results, which show a slight performance advantage for the CO-GAN, suggest that for this specific jamming detection problem, a small sacrifice in absolute distributional fidelity in favor of enhanced class separability is a beneficial trade-off.

However, this also highlights a potential risk: in scenarios with more complex or heavily overlapping classes, a task-oriented generator might be more susceptible to issues like mode collapse, whereas a fidelity-focused generator could prove to be more robust and generalizable.

3.2.4 Tailoring GANSec for Wireless Data

Generating realistic synthetic wireless data requires specific adaptations beyond standard GAN implementations, as naive approaches often fail to capture the unique characteristics of wireless signals. GANSec incorporates several critical design considerations to handle these intricacies effectively.

- **Handling Temporal Dependencies:** Wireless signal measurements are time-series data

Algorithm 3 Classification Oriented GAN Training: $\mathcal{D}$ performs a dual role: it assesses the realism of the input sample and predicts its class label.

Require: Feature set $x \sim \text{dist}_{\{y,c\}}(x)$ with label y and condition c

Ensure: Discriminator $\mathcal{D}$, Generator $\mathcal{G}$

- 1: Initialize number of epochs ϵ , loss functions $\mathcal{L}_i$, loss weights ω_i
- 2: Concatenate $\{y, c\}$
- 3: **for** current_epoch $\leftarrow 1, 2, \dots, \epsilon$ **do**
- 4: Sample noise $z \sim \mathcal{N}(0, I)$
- 5: Generate synthetic data using $\mathcal{G}$: $\hat{x} \leftarrow \mathcal{G}(z, \{y, c\})$
- 6: Discriminate real or synthetic data and predict label using $\mathcal{D}$

$$\{\text{output_real}, y'\} \leftarrow \mathcal{D}(x, c)$$

$$\{\text{output_synthetic}, \hat{y}\} \leftarrow \mathcal{D}(\hat{x}, c)$$

- 7: Compute loss for $\mathcal{G}$ and $\mathcal{D}$

$$\text{loss}_{\mathcal{D}} \leftarrow \omega_0 \times \mathcal{L}_0(1, \text{output_real}) + \omega_1 \times \mathcal{L}_1(y, y') +$$

$$\omega_2 \times \mathcal{L}_0(0, \text{output_synthetic}) + \omega_3 \times \mathcal{L}_1(y, \hat{y})$$

$$\text{loss}_{\mathcal{G}} \leftarrow \omega_4 \times \mathcal{L}_0(1, \text{output_synthetic}) + \omega_5 \times \mathcal{L}_1(y, \hat{y})$$

- 8: **end for**
-

with strong temporal correlations that are crucial for context. To capture these, GANSec employs backbones suited for sequential data. LSTMs are used for their ability to model long-range dependencies, while 2D CNNs effectively learn local, time-invariant patterns within the signal sequences. An MLP baseline is included for comparison, which requires flattening the time-series segments into 1D vectors, a process that can obscure important temporal relationships.

- **Preserving Statistical Properties:** Beyond temporal structure, wireless signals possess specific statistical properties that are vital for accurate representation. For instance, metrics like RSRP are often measured on a logarithmic scale (dBm). This logarithmic nature can interact suboptimally with standard neural network activation functions, and common loss functions behave differently in the logarithmic domain compared to the linear power domain. GANSec addresses this implicitly. While it does not apply specialized preprocessing, its generator must learn the complex, non-linear transformations required to account for these statistical intricacies. The discriminator's adversarial evaluation pushes the generated data towards realistic distributions, effectively handling the challenges posed by logarithmic scales within the generative framework.
- **Incorporating Domain Knowledge:** The conditional nature of GANSec allows for the explicit guidance of the generation process using domain-specific information. The condition vector can include parameters representing known environmental factors, device types,

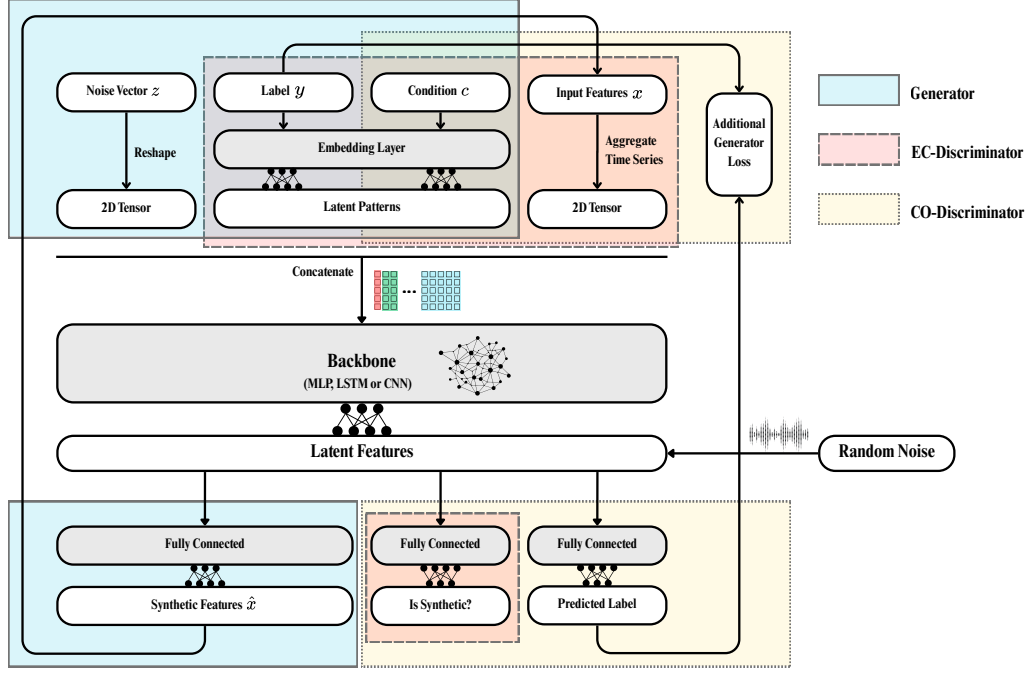


Figure 3.2: EC- and CO-training mechanisms within GANSec framework.

protocol states, or specific interference signatures, enabling the generation of contextually relevant and realistic data.

3.3 Evaluation of the Synthetic-Only Training Strategy

The validation of the GANSec framework is conducted on a comprehensive, multi-faceted evaluation designed to rigorously assess its ability to enhance detector robustness. The evaluation methodology moves beyond simple accuracy metrics to include qualitative analysis of the generated data, quantitative measures of its statistical properties, and a series of ablation studies to justify the framework's core design choices. The central pillar of this evaluation is a challenging cross-scenario experiment that directly tests the primary hypothesis: that a detector trained exclusively on synthetic data can achieve superior generalization to unseen environments.

3.3.1 Experimental Design and Baselines

The experimental setup is designed to simulate a realistic deployment scenario where a model trained in one environment must perform effectively in another. To achieve this, two distinct datasets were collected using SHIELD pipeline from our realistic 5G indoor testbed as described in previous Section 2.3.

- **Scenario A (Training Domain):** This dataset was used for training the generative models (GANSec and baselines). It consists of 3000 labeled time-series samples of 5G signal quality indicators, with jamming anomalies representing 46.37% of the data.
- **Scenario B (Testing Domain):** This dataset serves as the unseen test set to evaluate generalization. It was collected from a different physical location within the same environment, with an additional obstacle placed between the transmitter and receiver to induce a significant distribution shift. It contains 4559 samples, with a different class balance where anomalies constitute 65.08% of the data.

A key aspect of our proposed methodology involves training the downstream anomaly detection model exclusively on the synthetic data. Such training strategy offers several significant advantages compared to using the original real-world data:

1. It allows for the generation of perfectly balanced datasets, mitigating issues caused by original class imbalance.
2. If the GAN model captures the underlying distribution well, the synthetic data can exhibit greater diversity than the initial samples, potentially enhancing detector robustness and generalization.
3. It offers potential privacy preservation benefits, as training the anomaly detector solely on synthetic data avoids direct exposure of potentially sensitive real-world measurements.

Once GANSec is trained on the available real data using either the EC or CO mechanism, we utilize the trained $\mathcal{G}$ to produce synthetic data. To generate synthetic samples corresponding to a specific class label and condition $\{y, c\}$, we sample a random noise vector z and provide the tuple $\{z, y, c\}$ as input to the trained $\mathcal{G}$. This process, detailed in Algorithm 4, allows us to generate any desired number of synthetic samples for each class, potentially including optional post-processing or validation steps (e.g., using the $\mathcal{D}$ to filter generated samples based on a quality threshold t). We applied a quality threshold of $t = 0.5$ for the subsequent evaluation.

The downstream anomaly detection model is an LSTM-based classifier designed to classify wireless measurements as 'normal' or 'anomalous'. It is chosen for its proficiency in handling time-series data. The architecture consists of a 3-layer Long Short-Term Memory (LSTM) network, with each layer containing 128 hidden units, followed by a dropout layer with a rate of 0.3 to prevent overfitting. The final classification is performed by a fully connected dense layer with a sigmoid activation function. The model was trained for 256 epochs with a batch size of 256, using the Adam optimizer with a learning rate of 0.0001 and binary cross-entropy as the loss function. Its final performance is then measured on the real-world Scenario B dataset.

The performance of GANSec is benchmarked against three baseline data strategies:

- **Raw Data:** The detector is trained directly on the original, real data from Scenario A. This establishes the performance baseline and quantifies the negative impact of the distribution shift.

Algorithm 4 Synthetic Data Generation: $\mathcal{G}$ produces regulated synthetic feature sets given a target label $\hat{y}$ and condition $\hat{c}$, including an optional step for validating the quality of generated samples using the $\mathcal{D}$ with a threshold t

Require: Generator $\mathcal{G}$, Discriminator $\mathcal{D}$, label $\hat{y}$ and condition $\hat{c}$

Ensure: Synthetic feature set $\hat{x}$

- 1: Initialize regulations $regul(x)$, discriminator threshold t
 - 2: Concatenate $\{y, c\}$
 - 3: **while** $p < t$ **do**
 - 4: Sample noise $z \sim \mathcal{N}(0, I)$
 - 5: Generate synthetic data using $\mathcal{G}$: $\hat{x} \leftarrow \mathcal{G}(z, \{\hat{y}, \hat{c}\})$
 - 6: Regulate synthetic data: $\hat{x}' \leftarrow regul(\hat{x})$
 - 7: Validate using $\mathcal{D}$: $p \leftarrow \mathcal{D}(\hat{x}')$
 - 8: **end while**
-

- **Classical Oversampling:** The detector is trained on data from Scenario A augmented using DeepSMOTE [55], a state-of-the-art oversampling technique that fuses Autoencoder basics and SMOTE. This represents a strong baseline from the family of traditional, interpolation-based augmentation methods.
- **Unconditional GAN:** The detector is trained on data generated by TAnoGAN [30], an alternative GAN-based framework for time-series anomaly detection. This serves as a baseline to assess the specific contributions of the GANSec architecture and training objectives.

3.3.2 Training Settings for Diversity-Fidelity Balance

A fundamental challenge in training any generative model is managing the inherent trade-off between the fidelity (realism) and diversity (variety) of the generated samples [120, 119, 98]. For data augmentation, the goal is to generate samples that are both realistic enough to be useful for training and diverse enough to improve generalization. The GANSec training process was carefully configured to navigate this trade-off.

A key hyperparameter adjustment involved pruning the discriminator $\mathcal{D}$'s learning rate. During initial experiments, we observed that the $\mathcal{D}$ often converged much faster than the generator $\mathcal{G}$. A $\mathcal{D}$ that is too powerful provides uninformative, vanishing gradients to the $\mathcal{G}$, effectively halting the $\mathcal{G}$'s learning process. By reducing the $\mathcal{D}$'s learning rate relative to the $\mathcal{G}$'s, a more stable training dynamic was achieved, allowing the $\mathcal{G}$ to progressively improve without being overwhelmed. For instance, when training the MLP-based GAN, we used the following hyperparameters:

- Number of epochs: 5000
- Batch size: 128
- Generator learn rate: 0.00012

- Discriminator learn rate 0.0001
- Loss function: Binary Cross Entropy
- Optimizer: Adam

The learning curves for the EC-GAN and CO-GAN with the MLP backbone, shown in Figure 3.3a and Figure 3.3b respectively, illustrate the results of these settings. Both plots show a healthy competitive learning process, where the loss values for the $\mathcal{G}$ and $\mathcal{D}$ fluctuate over the 5000 epochs, indicating a stable adversarial dynamic rather than one model overpowering the other. The evolution of the synthetic samples, displayed at the top of each plot, visually confirms the $\mathcal{G}$'s learning trajectory. At the initial stage (epoch 1), the $\mathcal{G}$ outputs pure noise. As the $\mathcal{D}$ learns the original data distribution, the $\mathcal{G}$ begins to produce high-fidelity samples with minor noise in the middle stage (around 1000-2000 epochs). In the later stages, the $\mathcal{G}$ competes with a proficient discriminator, continuously refining its output to find a balance between fidelity and diversity.

A comparison of the learning curves reveals that the EC discriminator's loss is more stable, likely due to its simpler, single-task objective of authenticity classification. The CO discriminator's loss exhibits more fluctuation, reflecting the increased complexity of its dual task: assessing authenticity while also performing class prediction.

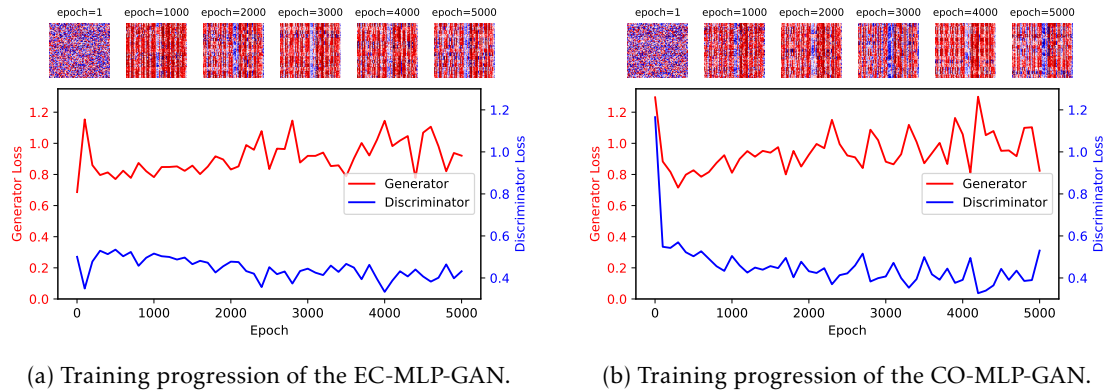


Figure 3.3: Training dynamics of the EC-GAN and CO-GAN. Both plots show the loss over 5000 epochs. Above each plot are examples of generated samples at key epochs.

3.3.3 Quantitative Analysis of Synthetic Data Quality

To quantify the similarity between the generated and real data distributions, Fréchet Inception Distance (FID) and Maximum Mean Discrepancy (MMD) are employed. Lower scores on these metrics indicate a higher degree of similarity to the original data distribution (Scenario A). The results, however, must be interpreted with nuance, considering the trade-off between fidelity and

diversity. As shown in the Cumulative Distribution Function (CDF) plots in Figure 3.4 and 3.5, the different augmentation methods produce synthetic data with distinct statistical profiles.

The analysis of FID scores, which measure the distance between feature distributions of real and generated samples, reveals a clear hierarchy among the methods. The DeepSMOTE baseline consistently achieves the lowest FID scores (average 1.603), with its CDF rising almost vertically near zero. This indicates that its generated samples are statistically very close to the original data, reflecting high fidelity. However, this is an expected outcome of its interpolative nature, which inherently limits the generation of novel, diverse samples beyond the convex hull of the original data.

In contrast, the GANSec variants exhibit higher and more broadly distributed FID scores, with values for the CNN backbones reaching up to 60-70. Critically, in the context of training for robustness, these higher scores are not interpreted as a failure but as a sign of beneficial diversity. The goal is to train a detector that can generalize to an unseen distribution (Scenario B). To achieve this, the training data must be more diverse than the limited samples available in Scenario A. A generator that successfully produces a more diverse set of samples will naturally exhibit a greater statistical distance from the limited original dataset. Therefore, the higher FID scores for GANSec, when viewed in conjunction with its superior downstream task performance, suggest that the framework is achieving a beneficial increase in diversity that is crucial for generalization, rather than suffering from a simple lack of fidelity.

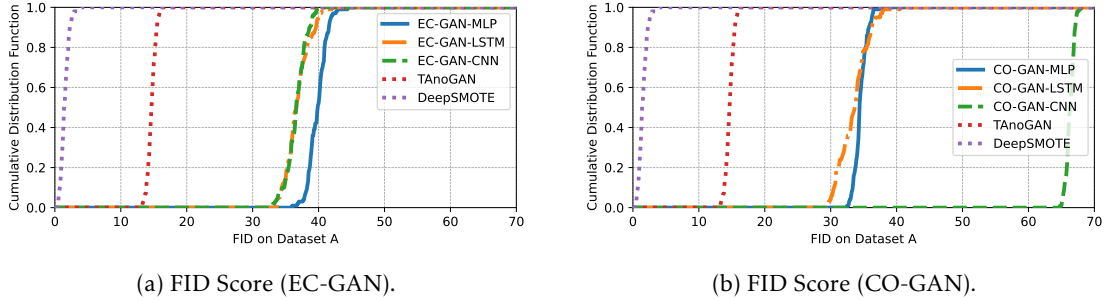


Figure 3.4: CDF of FID Scores for Synthetic Samples over 100 Generation Iterations.

A similar pattern emerges from the MMD scores, which also quantify the discrepancy between two distributions. DeepSMOTE again shows the lowest scores (average 0.164), confirming its high-fidelity but low-diversity output. The TAnoGAN baseline presents an interesting case, with a low FID (14.762) but a high MMD (0.632), suggesting it captures the global data structure but may struggle with local sample quality or mode diversity.

The GANSec variants generally have higher MMD scores than DeepSMOTE, reinforcing the observation that they generate a more varied set of samples. The MLP-based GANs tend to have lower MMD scores compared to their LSTM and CNN counterparts, indicating less diversity, which aligns with their weaker performance on the downstream task. The overall analysis of both FID and MMD supports the central hypothesis: the "aggressiveness" of the CNN-based

CNNs, reflected in their higher dissimilarity scores, produces a more diverse synthetic dataset. This increased diversity, while technically representing lower fidelity to the original training set, is precisely what equips the downstream model with the robustness needed to generalize effectively to the unseen test set.

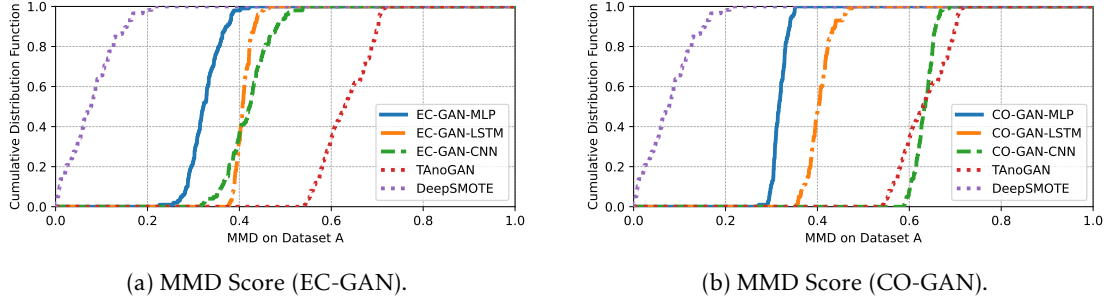


Figure 3.5: CDF of MMD Scores for Synthetic Samples over 100 Generation Iterations.

3.3.4 Cross-Scenario Evaluation of Anomaly Detector Robustness

The central hypothesis of this research is that an anomaly detector trained exclusively on data generated by the GANSec framework can achieve superior robustness and generalization capabilities, particularly when faced with a distribution shift between training and testing environments. The cross-scenario evaluation, where a model is trained on data from Scenario A and tested on the unseen Scenario B, was designed to directly validate this claim. The comparative performance of various methods is illustrated in the accompanying Figure 3.6 and summarized in Table 3.4.

Model	Aug. Ratio	Acc.	Prec.	Rec.	F1	ROC.
Raw Data (<u>Baseline</u>)	N/A	0.781	0.914	0.735	0.815	0.802
DeepSMOTE	+200%	0.822	0.765	0.794	0.779	0.844
TAnoGAN	+500%	0.833	0.928	0.806	0.863	0.845
EC-GAN-MLP	+30%	0.851	0.836	0.981	0.903	0.792
EC-GAN-LSTM	+150%	0.910	0.962	0.922	0.942	0.905
EC-GAN-CNN	+150%	0.883	0.925	0.894	0.910	0.878
CO-GAN-MLP	+20%	0.836	0.823	0.975	0.893	0.772
CO-GAN-LSTM	+300%	0.894	0.936	0.884	0.910	0.922
CO-GAN-CNN	+70%	0.921	0.945	0.934	0.940	0.915

Table 3.4: Peak Performance Metrics on Unseen Scenario B. For each model, the table reports key metrics at the augmentation ratio that yielded its optimal accuracy.

First, the severity of the generalization challenge is established by the Raw Data baseline. As depicted by the flat red dashed line in both benchmark comparison figures, the detector trained

solely on real data from Scenario A achieves a mere 78.1% accuracy when tested on Scenario B. This significant performance drop from its in-domain accuracy quantifies the performance collapse caused by the distribution shift and serves as the primary benchmark for all other methods.

The conventional augmentation baselines, DeepSMOTE and TAnoGAN, show only modest improvements. The detector trained with DeepSMOTE reaches a peak accuracy of approximately 82.2%, while the TAnoGAN-trained model achieves a slightly better 83.3%. While these results indicate that both interpolative and basic generative augmentation provide some benefit, their performance remains limited. Furthermore, as shown in the figures, the accuracy of these methods is often volatile, fluctuating significantly with the volume of augmented data, which suggests a lack of stable generalization. Ultimately, they are insufficient to fully bridge the generalization gap.

In stark contrast, GANSec variants demonstrate a profound and consistent improvement in robustness. The results from both the EC and CO-GAN configurations validate the effectiveness of the synthetic-only training strategy.

- **CO-GAN Performance:** As illustrated in Figure 3.6b, the CNN configuration (solid green line) consistently outperforms all other methods across nearly the entire spectrum of data augmentation ratios. It achieves a peak accuracy of 92.1% at a +70% augmentation volume, effectively erasing the performance degradation caused by the distribution shift. The LSTM model also shows exceptionally strong performance, rivaling and even surpassing the CNN variant at higher augmentation ratios.
- **EC-GAN Performance:** Figure 3.6a shows that this success is not unique to one configuration. The EC-GAN variants also decisively outperform the baselines. Here, the LSTM model achieves a peak accuracy of over 91% at a +150% ratio, reinforcing the framework's overall strength.

The success of these models is remarkable for two reasons. First, they completely mitigate the performance collapse observed in the baseline. Second, they achieve a level of accuracy on an unseen domain that is comparable to, or even exceeds, what a model trained on real data might achieve within its own domain. The high F1-score (0.940), precision (0.945), and recall (0.934) for the top-performing CO-GAN-CNN model further confirm that this improvement is well-balanced and not skewed towards a particular type of error.

Collectively, these results provide strong empirical validation for our primary hypothesis, demonstrating that the tailored GANSec framework is highly effective for training robust, generalizable detectors using a synthetic-only data strategy.

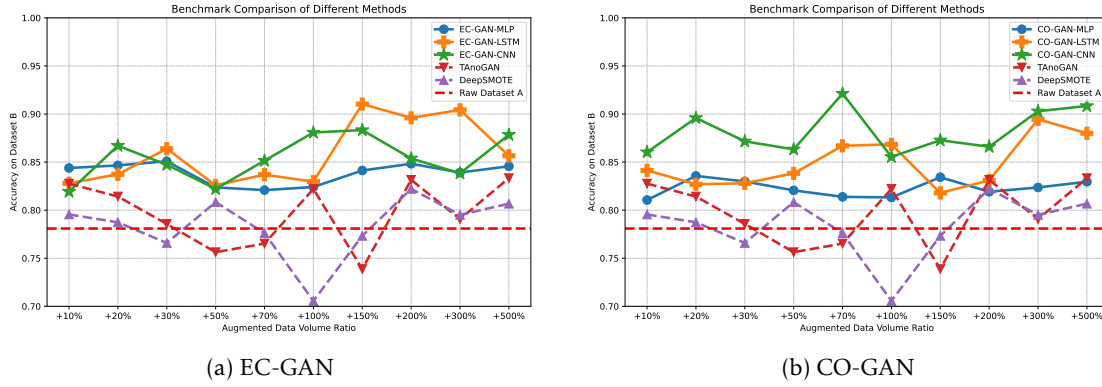


Figure 3.6: Impact of the augmentation ratio (relative size of synthetic data used for training) on the accuracy achieved on the unseen Dataset B.

3.4 Chapter Summary

This chapter has systematically addressed the critical data-centric challenges that impeded the development of robust wireless anomaly detection systems. It began by establishing a deep motivation, framing the problem not as a simple lack of data, but as a confluence of a severe data acquisition bottleneck and the statistical inevitability of generalization failure. The non-stationary nature of wireless environments continuously induces distribution shifts, which fundamentally violate the core assumptions of supervised learning models and render traditional augmentation techniques based on linear interpolation theoretically insufficient.

In response to this challenge, this work proposed and detailed GANSec, a conditional GAN framework tailored for generating high-fidelity wireless time-series data. The framework's design was thoroughly justified, including its exploration of sequence-aware neural network backbones (MLP, LSTM, CNN) and its novel investigation into two distinct training objectives: a fidelity-focused Embedded Conditional (EC) approach and a task-oriented Classification Oriented (CO) approach.

The efficacy of GANSec was validated through a comprehensive evaluation centered on a synthetic-only training strategy. This multi-faceted assessment provided compelling evidence for the framework's success. Quantitative analysis of data quality metrics contextualized the diversity-fidelity trade-off, arguing that increased diversity, even at the cost of perfect fidelity to the training set, is beneficial for robustness. Finally, the primary hypothesis was validated in a challenging cross-scenario experiment, where a detector trained exclusively on GANSec's synthetic data achieved an accuracy of 92.1%, dramatically outperforming all baselines and demonstrating exceptional generalization.

In conclusion, this chapter positions tailored generative data augmentation as a foundational and highly effective methodology for overcoming the data-centric limitations in wireless security. The success of the synthetic-only training strategy demonstrates that it is possible to create

detectors that are not only resilient to the performance degradation caused by environmental changes but are, in fact, more robust and generalizable than models trained on limited real-world data alone.

A Framework for Collaborative Intelligence

The preceding chapters of this dissertation have established the design and validation of a robust on-device jamming detector. This detector, operating at the individual UE level, demonstrates high accuracy in identifying jamming conditions using native signal quality metrics, as evidenced by the performance of machine learning models achieving F1-scores up to 92.3% under experimental conditions. However, the utility of such a detector, when operating in isolation, is fundamentally constrained. Its perception is confined to a single point in the network, rendering it incapable of assessing the broader context of a detected anomaly. This inherent lack of spatial awareness is a critical deficiency for understanding and responding to network-wide threats.

This chapter addresses this limitation by proposing a conceptual framework that transcends the single-node paradigm. It introduces a system designed to scale from individual sensing to collective, network-wide intelligence, thereby transforming localized data points into actionable situational awareness.

4.1 Fundamental Limitations of On-Device Threat Detection

The efficacy of security systems is predicated on the quality and context of the information it receives. For an isolated detector operating at the physical layer of a wireless network, the available information is severely constrained, creating an informational vacuum that fundamentally limits its ability to provide actionable intelligence.

This section systematically deconstructs the critical weaknesses inherent in on-device detection, demonstrating why a collaborative approach is not merely an improvement but an absolute necessity for robust network security.

4.1.1 Inability to Assess Geographic Scope and Scale

A primary and critical failure of an isolated detector is its inherent inability to assess the geographic scope and scale of a perceived threat. A single sensor's measurement cannot reveal the whole characteristics of a larger spatiotemporal field. This creates a fundamental spatial ambiguity. From the perspective of a single device, a powerful jamming signal caused by congestion in a crowded channel is indistinguishable from a sophisticated, wide-area jamming attack targeting an entire metropolitan area [267]. The local observation is identical in both scenarios, yet the implications for network integrity are vastly different.

From a research perspective, this is a problem of insufficient observability and incomplete information. An isolated node has no knowledge of the spatial gradients or correlations of the phenomenon it is observing. For example, the power of a jamming signal decays with distance, an isolated node measures only a single point on this decay curve, making it impossible to infer the source, power, or area of effect [45, 171]. Without this contextual data from distributed points, any detected event is underdetermined, which prevents the accurate modeling of the threat's physical extent or severity [122]. This limitation is a significant obstacle to the development of robust, autonomous security protocols, which require the ability to differentiate between systemic threats and localized, uncorrelated faults to function effectively.

The challenge is magnified exponentially in the context of modern network paradigms like the Space Information Networks (SIN) [112]. The attack surface in these environments is vast and radically heterogeneous, comprising billions of interconnected devices. A single smart thermostat reporting anomalous RF interference provides no actionable intelligence as to whether the threat is confined to one smart home or is impacting the critical infrastructure of an entire smart city [166]. A collaborative framework transforms these isolated reports into a coherent map of the threat's footprint, enabling a targeted and proportional response.

4.1.2 The Inherent Incapability of Single-Point Source Localization

Pinpointing the physical origin of a radio frequency signal is a process that, by the laws of physics, requires measurements from multiple, spatially distributed vantage points. An isolated device can report that it is experiencing interference, but it can provide no information regarding the direction or location of the interference source. This leaves security and law enforcement teams effectively blind, unable to dispatch response teams, implement geographically targeted countermeasures, or hold malicious actors accountable.

Common RF source localization techniques fundamentally rely on the principle of triangulation or trilateration, which requires a distributed network of sensors. These techniques include:

- **Time Difference of Arrival (TDoA):** This method measures the difference in the arrival time of a signal at multiple receivers. Since the signal propagates at a known speed, these

time differences define a set of hyperboloids on which the source must lie. The intersection of these hyperboloids reveals the source's location [228, 194].

- **Power Difference of Arrival (PDoA):** This technique compares the received signal strength at different sensors. Based on a known path-loss model, the differences in power can be translated into differences in distance, which can be used to estimate the source's location [228].
- **Angle of Arrival (AoA):** This approach requires sensors equipped with antenna arrays to determine the direction from which a signal is arriving. The intersection of the directional bearings from multiple sensors points to the location of the transmitter [183, 129].

All of these methods are impossible to implement with a single, isolated sensor [78]. This limitation is particularly acute for the billions of low-power, resource-constrained devices that constitute the IoT and WSN. These devices often lack the computational power, energy budget, or sophisticated hardware (like antenna arrays) required for advanced on-device signal processing [253, 39]. For these networks, collaboration is not merely beneficial for localization, but the only viable path to achieving it.

4.1.3 Resolving Ambiguity: Differentiating Malicious Attacks from Benign Faults

One of the most challenging tasks for an isolated detector is resolving the ambiguity between a malicious attack, a benign hardware or software fault, and unintentional interference. A single UE reporting a service disruption could be the victim of a targeted attack, or it could simply be experiencing a hardware failure, a software bug, or interference from a nearby, non-malicious device.

A collaborative framework resolves this ambiguity through spatiotemporal pattern analysis. By aggregating reports from across the network, the system can identify correlations that betray the nature of the underlying event. For example: a geographically coherent cluster of UE all reporting a sudden and simultaneous drop in signal quality is a strong indicator of a singular, powerful jamming source operating in that area [270, 23]. Conversely, a series of sporadic, uncorrelated reports of service disruption occurring at different times and in different locations across the network is far more likely to be the result of random, independent device faults.

This principle of pattern analysis is also critical for detecting spoofing attacks. A key challenge is distinguishing a legitimate user moving through the network from a spoofer initiating a transmission from a different physical location. An isolated access point would see a change in the RSS in both scenarios. However, a collaborative network of access points can analyze the temporal evolution of the RSS signature across the entire sensor array. The movement of a legitimate user creates a smooth and continuous pattern of change in the RSS readings. In contrast, a spoofing attack, where the transmission source instantly switches from the legitimate

user's location to the attacker's, creates an abrupt, discontinuous jump in the spatial signature of the RSS readings [87, 201]. This spatiotemporal correlation, which is impossible for an isolated device to perform, provides a robust, non-cryptographic method for unmasking the impersonator [138].

Therefore, the move to collaborative intelligence is not merely an incremental improvement but an essential and transformative step. It is predicated on the principle of emergent properties: by aggregating a multitude of simple, distributed observations, the system can achieve true situational awareness that is impossible for any individual component to attain. This collective sense-making enables network operators to shift from a purely reactive posture, where they can only confirm that a single device is experiencing a service disruption, to a proactive and strategic one, where they can comprehend the threat's scope, estimate its location, and infer its characteristics.

4.2 A Lightweight Collaborative Intelligence Architecture

The proposed framework is architected to translate the theoretical benefits of collaboration into a practical system for 5G and beyond, including massive IoT deployments. Its design is guided by a set of foundational principles aimed at ensuring feasibility, scalability, and robustness in real-world environments.

4.2.1 Core Architectural Principles

The architecture is founded on three core principles: minimalism, scalability, and decentralized sensing. Each principle represents a deliberate design choice to balance performance with the practical constraints of modern wireless networks.

- **Minimalism:** This principle dictates that the burden on individual UE must be as low as possible. For a collaborative system to be viable across a vast and heterogeneous population of devices, many of which are resource-constrained IoT sensors, the on-device requirements for processing, power consumption, and data transmission must be negligible. The framework achieves this by limiting the UE's role to running a lightweight local detector and reporting a single bit of information, thereby maximizing the potential for widespread participation.
- **Scalability:** A direct consequence of minimalism, scalability is the architecture's capacity to function effectively as the number of participating UE grows into the millions or billions. By relying on minimalist binary reports, the central fusion entity is protected from being overwhelmed by a high volume of complex data. The processing load of aggregating binary inputs is significantly lower than that of fusing rich feature vectors, enabling the system to scale gracefully with network density.

- **Decentralized Sensing:** The framework employs a strategic division of labor. The task of sensing, the initial detection of a jamming anomaly, is fully decentralized and pushed to the network edge, leveraging the sensing capabilities of every participating UE. The more computationally intensive task such as aggregating data, fusing evidence, and localizing the source, is handled by a central or distributed fusion entity with greater computational resources. This architectural pattern efficiently offloads complexity from the constrained edge devices to the more powerful core or edge cloud.

A critical design choice underpinning these principles is the use of binary reporting, which presents a fundamental trade-off between information fidelity and system efficiency. This is a form of 'hard decision fusion', where each UE quantizes its local observation into a definitive 'JAMMED' or 'NOT_JAMMED' state before transmission. The primary disadvantage of this approach is the inherent information loss. A continuous metric, such as the SINR, contains far more nuance than a binary flag. UE experiencing a marginal drop in SINR might be forced to report 'NOT_JAMMED', while one just below the threshold reports 'JAMMED', even though their physical conditions are very similar. This loss of granularity can make individual reports more susceptible to false positives and negatives, particularly for UE in ambiguous signal conditions near the edge of a jamming-affected area.

However, this trade-off is made deliberately in favor of achieving massive scalability and operational simplicity. The advantages are significant. First, the communication overhead per device is reduced to its theoretical minimum: a single status bit, plus location and identity metadata. This is essential for energy-constrained devices and for preventing network congestion. Second, this simplicity may confer a degree of robustness. Binary decisions can be less sensitive to minor sensor calibration drifts or low-level noise compared to continuous values. The core philosophy of the framework is that the information lost at the individual level is robustly recovered at the collective level. The sheer quantity and spatial diversity of thousands or millions of simple binary reports provide a rich dataset from which a high-fidelity, network-wide understanding can be reconstructed.

4.2.2 System Components and Workflow

The operational flow of the architecture, as illustrated in Figure 4.1, the conceptual diagram, proceeds from distributed sensing to centralized intelligence generation. The system comprises three primary components.

- **On-Device Sensing Agent:** This is a lightweight software module residing on each participating UE. Its responsibilities are strictly limited to periodically analyzing local 5G signal quality measurements (e.g., RSRP, RSRQ, SINR) accessible through standard device interfaces, applying the pre-trained anomaly detection model to generate a binary decision, and packaging this decision for transmission.

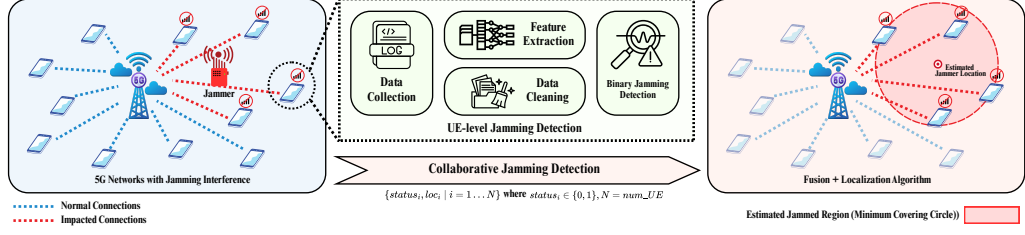


Figure 4.1: Overview of the proposed collaborative jamming detection workflow, progressing from distributed UE-level sensing and binary decision-making to centralized fusion and jammer localization.

Secure Reporting Channel: This component defines the communication pathway for transmitting the data packet from the UE to the fusion entity. A typical packet would contain the UE identifier, a timestamp, its current location, and the binary jamming status. The security of this channel is paramount to prevent data tampering or spoofing.

Central/Distributed Fusion Entity: This is a logical entity, which could be implemented as a cloud-based service, a NFV at the mobile edge, or a distributed system. Its sole purpose is to receive and aggregate the reports from all UE and execute the core fusion and localization algorithms.

The end-to-end workflow unfolds in a continuous cycle:

1. UE across the network independently and concurrently monitor their local channel conditions.
2. Based on this monitoring, each UE makes a local, binary determination of its state: 'JAMMED' (represented as '1') or 'NOT_JAMMED' (represented as '0').
3. Each UE securely transmits its status report, timestamp, and location to the fusion entity. This reporting can be periodic or event-triggered to optimize overhead.
4. The fusion entity aggregates the continuous stream of incoming reports, using their spatial and temporal information to construct a real-time map of jamming detections.
5. The entity applies evidence fusion algorithms to this raw map to assess confidence and filter out noise, generating a high-fidelity Jammed Area Map.
6. Finally, it applies geometric localization algorithms to the Jammed Area Map to estimate the physical location of the jamming source.
7. The ultimate output is a comprehensive, network-wide intelligence report that is continually updated, providing operators with a dynamic view of the physical-layer threat landscape.

4.3 Fusion and Localization from Distributed Binary Data

The effectiveness of the framework hinges on its ability to transform a massive volume of simple, unreliable binary inputs into a coherent and accurate network-wide assessment. This is accomplished through a pipeline of specialized algorithms designed explicitly for this data context: an evidence-theoretic approach for robust data fusion and a boundary-based geometric method for localization.

4.3.1 Aggregating Unreliable Inputs with Evidence Theory

To handle the inherent uncertainty and potential unreliability of individual UE reports, the framework proposes the use of DST for data fusion. DST [210], also known as the theory of evidence, is a powerful mathematical framework that generalizes Bayesian probability, making it exceptionally well-suited for combining evidence from disparate, uncertain, and potentially conflicting sources [90, 124, 178].

Conceptually, DST operates by assigning belief mass not just to individual hypotheses but to sets of hypotheses. For this problem, the Frame of Discernment, denoted by Θ , is the set of all possible, mutually exclusive states for a given location: $\Theta = \{\text{JAMMED}, \text{NOT_JAMMED}\}$. Each UE report is treated as an independent piece of evidence that provides a Basic Probability Assignment (BPA), or mass function m , which distributes a total belief of 1 across the power set of Θ . For instance, a report of 'JAMMED' from a UE with a known reliability of 80% might be translated into a mass function such as:

- $m(\{\text{JAMMED}\}) = 0.8$ (the belief committed solely to the JAMMED state)
- $m(\{\text{NOT_JAMMED}\}) = 0.0$ (no belief committed to the NOT_JAMMED state)
- $m(\Theta) = 0.2$ (the remaining belief, which is uncommitted and represents uncertainty or ignorance)

This ability to explicitly model uncertainty is a key advantage of DST [124]. From the mass functions, two metrics are derived: Belief ($Bel()$) and Plausibility ($Pl()$). $Bel(A)$ is the sum of masses for all subsets of a hypothesis A , representing the minimum level of confidence in A . $Pl(A)$ is the sum of masses for all sets that intersect with A , representing the maximum possible confidence in A [178].

When multiple reports concerning the same location are received, they are combined using Dempster's rule of combination. This rule provides a mathematical method for merging the mass functions from two or more sources to produce a new, combined mass function that reflects the consensus of the evidence.

This approach is significantly more robust than simpler aggregation methods like majority voting. A majority vote is a "winner-takes-all" mechanism that treats every report as an equal, binary vote, completely abstracting away any notion of confidence or uncertainty [152, 93]. In

such a system, five low-confidence (and potentially erroneous) reports could outweigh four high-confidence reports. In contrast, DST's combination rule naturally reinforces belief where evidence is consistent and quantifies the degree of conflict where it is not. This allows the system to gracefully handle a certain percentage of noisy or malicious inputs, as conflicting evidence will reduce the combined belief rather than simply being outvoted. This is particularly crucial for reports from UE near the jamming boundary, where local conditions are ambiguous and uncertainty is high.

The process reveals a clear data-refinement pipeline: unreliable, individual binary decisions are first elevated into more nuanced evidence through the BPA. This evidence is then aggregated via Dempster's rule to produce a reliable spatial map of belief. This map, which represents a filtered and consolidated view of the network state, becomes the input for the subsequent localization stage. As a computationally lighter alternative for massive-scale deployments, the Ordered Counting Rule could also be investigated, which prioritizes reports from potentially more reliable UE to reach a decision faster.

4.3.2 Jammer Localization using Boundary-Based Techniques

Given that the input from the fusion stage is a spatial map of binary or high-belief states, traditional jammer localization techniques that rely on continuous physical-layer metrics like Time of Arrival (ToA) or RSSI are inapplicable. The problem must therefore be reframed as a geometric one: estimating the center of a region defined by a set of discrete points.

The first and most critical step in this process is the identification of 'boundary UE'. These are the devices that geographically delineate the transition between the 'JAMMED' and 'NOT_JAMMED' areas of the network [46, 245]. From the fused data map, these UE can be identified algorithmically, for instance, by selecting 'NOT_JAMMED' UE that have at least one 'JAMMED' UE within a defined neighborhood radius. These boundary points are the most informative data available, as they collectively form a perimeter that approximates the outer edge of the jammer's effective range [244].

With the set of affected UE identified, the framework proposes using a Minimum Enclosing Circle (MEC) algorithm for localization. The MEC is a classic computational geometry problem that seeks to find the unique circle of smallest radius that contains a given set of points [266]. The operational workflow is as follows: the fusion entity takes the set of coordinates of all UE within the high-confidence 'JAMMED' region and computes their minimum covering circle. The center of this circle is then designated as the estimated location of the jammer. An efficient approximation of the MEC algorithm suitable for this purpose is detailed in Algorithm 5.

The precision of this boundary-based approach is, however, subject to inherent limitations that are directly coupled to the density and spatial distribution of the participating UE.

- In a dense urban environment, a high concentration of UE will provide numerous data points near the true physical boundary of the jamming effect. This will result in a well-

Algorithm 5 Binary-Input Jammer Localization Using Minimum Enclosing Circle

Require: Set of UE reports $R = \{(status_i, loc_i) \mid i = 1 \dots N\}$, where $status_i \in \{0, 1\}$ (0: Not Jammed, 1: Jammed) and loc_i is UE location (e.g., coordinates).

Ensure: Estimated Jammer Location L_{jam} .

1: $P \leftarrow \{loc_j \in R \mid status_j = 1\}$

2: Find point $p_1 \in P$ with min(x-coordinate), point $p_2 \in P$ farthest from p_1 , and point $p_3 \in P$ farthest from p_2 .

$$p_2 \leftarrow \arg \max_{p \in P} Dist(p_1, p)$$

$$p_3 \leftarrow \arg \max_{p \in P} Dist(p_2, p)$$

3: Initialize Center point $c \leftarrow Mean(p_2, p_3)$

4: Initialize Radius $r \leftarrow 0.5 \times Dist(p_2, p_3)$

5: **for all** $p_i \in P$ **do**

6: $d_i \leftarrow Dist(p_i, c)$

7: **if** $d_i > r$ **then**

$$c \leftarrow c + \frac{(p_i - c)(d_i - r)}{2d_i}$$

$$r \leftarrow Mean(r, d_i)$$

8: **end if**

9: **end for**

10: **return** $L_{jam} \leftarrow c$

defined set of points for the MEC algorithm, leading to a high-confidence localization with relatively low error.

- Conversely, in a sparse rural area, the few available UE may be located far from the actual jamming boundary. The resulting set of points will be sparse and poorly distributed, leading to high geometric uncertainty. The computed minimum covering circle will be a coarse and potentially inaccurate approximation of the true jammed area, resulting in a large error in the estimated jammer location.

In the future, it may be possible to enhance accuracy by training Deep Learning models, such as spatial regression or graph neural networks, on simulated data [26]. Such models could learn the complex, non-linear mapping between a binary jam map and a jammer's location, potentially outperforming purely geometric methods, though at the cost of increased complexity and the need for extensive training data.

4.3.3 Deriving Fine-Grained Network Insights

The output of the collaborative framework extends beyond a mere location coordinate or binary detection status. By aggregating spatiotemporal data, it generates the Jammed Area Map, a real-time representation of the geographic extent of the attack. As a dynamic, network-wide

intelligence artifact, this map provides not only spatial localization but also enables the extraction of nuanced insights, as its geometric and temporal properties can be analyzed to infer critical characteristics of the jammer.

Inferring Jammer Characteristics from Map Shape

The shape of the high-confidence jammed area can provide strong clues about the jammer's antenna technology [69].

- An **omnidirectional jammer**, which radiates power equally in all directions, will produce a jamming effect that is roughly circular. If the Jammed Area Map consistently shows a nearly perfect circle of affected UE, it is a strong indicator of an omnidirectional antenna.
- A **directional jammer**, which focuses its power in a specific direction, will produce a non-circular pattern, such as a cone, wedge, or ellipse. The identification of such a shape in the Jammed Area Map would allow operators to infer not only the jammer's location but also its likely orientation.

Inferring Jammer Behavior from Temporal Analysis

The framework is designed to operate continuously, producing a sequence of Jammed Area Maps over time. Analyzing the evolution of this map provides powerful insights into the jammer's behavior.

- **Mobility Detection:** A static jammer will produce a stable map that does not change significantly over time. In contrast, a mobile jammer (e.g., a device in a moving vehicle) will cause the cluster of 'JAMMED' reports to translate across the map. By tracking the centroid of the jammed area at successive time intervals, the system can detect this movement, estimate the jammer's trajectory and approximate speed, and potentially predict its future path [223].
- **Behavior Pattern Recognition:** Different jamming strategies will manifest as distinct temporal patterns in the map [54]. A constant jammer produces a persistent, stable jammed area. A sweeping jammer, which varies its frequency or direction over time, might cause the jammed area to appear to pulsate, rotate, or change shape periodically. A reactive jammer, which transmits interference only when it detects legitimate network traffic, would cause the Jammed Area Map to appear and disappear intermittently, potentially in correlation with known network traffic patterns [204]. This level of insight allows for a more tailored and effective network response.

4.4 In-Depth Analysis of Operational Challenges and Mitigations

While the proposed framework offers a promising conceptual blueprint, its transition to a practical, deployed system necessitates addressing several significant operational challenges inherent to large-scale distributed sensing systems. A critical analysis of these challenges and an exploration of potential mitigation strategies, drawing from broader principles in distributed systems, wireless security, and data science, is essential.

Table 4.1 provides a high-level summary of the primary challenges, their potential impact on the system, and the classes of mitigation strategies that can be employed to address them.

4.4.1 Communication Overhead

While binary reporting is designed to be lightweight on a per-message basis, the aggregate traffic from millions of devices reporting periodically could still impose a substantial load on the network, leading to congestion and increased latency. For battery-powered IoT devices, frequent transmissions can also lead to premature energy depletion.

To mitigate this, the framework must move beyond simple periodic reporting and adopt adaptive reporting strategies [62]. The most straightforward approach is event-triggered reporting, where UE only transmits data when its detected status changes (e.g., from 'NOT_JAMMED' to 'JAMMED'). In static or slow-changing jamming scenarios, this would reduce network traffic by orders of magnitude. More sophisticated methods could include variance-triggered sampling, where reporting frequency increases only when local signal metrics become unstable, or model-based reporting, where UE transmits only when its local observation deviates significantly from a prediction made by the fusion center. Additionally, data compression techniques can be applied, particularly to the location metadata. For slow-moving UE, delta encoding can be used to transmit only the change in location from the last report. For static devices, techniques like Run-Length Encoding (RLE) can be used to compress long periods of unchanged status reports [226, 91].

4.4.2 Data Reliability and FDI Resilience

The framework's reliance on crowd data source makes it vulnerable to both unintentional errors (e.g., faulty sensors, transmission corruption) and malicious attacks, particularly False Data Injection (FDI) [151]. A coordinated FDI attack could be used to create "ghost" jammers, triggering false alarms and wasting operator resources, or to mask a real jammer by injecting false 'NOT_JAMMED' reports within an affected area.

A multi-layered defense is required to ensure data reliability. The first layer is cryptographic authentication. Each report should be digitally signed to verify its origin and integrity. Given

Challenge	Potential Impact	Proposed Mitigation Strategies
Communication Overhead	Network congestion, increased latency, and battery drain on UE in massive IoT scenarios.	Adaptive Reporting: Event-triggered (report on status change), variance-triggered, or model-based reporting [62]. Data Compression: Lightweight lossless (e.g., RLE) or lossy techniques for location data [53, 226].
Data Reliability & FDI Resilience	Corrupted global view, incorrect localization, DoS through false alarms.	Cryptographic Authentication: Lightweight authenticated encryption to verify report origin and integrity [233]. Anomaly Detection: Spatial consistency checks and ML models to detect anomalous report patterns [248, 170]. Trust Systems: Dynamically score UE reliability based on historical reporting behavior [212].
Impact of UE Density and Distribution	Severe degradation of localization accuracy and boundary mapping in sparse areas.	Predictive Modeling: Use physics-informed models to predict the jamming boundary, calibrated by sparse UE data [75, 260]. Data Prioritization: Assign higher weight in the fusion process to reports from UE near the estimated jamming boundary.
System Dynamics and UE Mobility	Stale data from mobile UE blurring the jamming boundary, leading to inaccurate maps.	Time Weighting: Assign decaying weights to older reports in the fusion algorithm. Age of Information (AoI): Use AoI as a metric to explicitly manage and minimize data freshness, ensuring the global view is timely [215, 148].
Handling Multiple Jamming Sources	MEC algorithm fails, producing a single, incorrect location estimate for multiple jammers.	Spatial Clustering: Employ algorithms like DBSCAN [68] as a preprocessing step to partition 'JAMMED' reports into distinct clusters before applying MEC to each cluster individually [214].

Table 4.1: Summary of Operational Challenges and Mitigation Strategies.

the resource constraints of many IoT devices, this necessitates the use of lightweight cryptography. Authenticated encryption schemes built on efficient primitives like TWINE [229] or PRESENT [34] are designed for small footprints and low computational complexity, making them feasible for securing reports from even the most constrained devices [233]. The second layer involves anomaly detection at the fusion entity. Beyond the spatial consistency checks inherent in the DST fusion process, dedicated ML models can be trained to detect anomalous patterns in the report stream itself, such as a sudden, statistically unlikely burst of 'JAMMED' reports from a single geographic area [248, 170]. The third layer is the implementation of a

trust and reputation system [212]. Each participating UE would be assigned a dynamic trust score. This score would be increased when its reports are consistent with those of its neighbors and decreased when they are in conflict. Reports from UE with low trust scores would be down-weighted or discarded by the fusion algorithm, effectively marginalizing unreliable or malicious nodes [44].

4.4.3 Impact of UE Density and Distribution

The accuracy of both the Jammed Area Map and the MEC-based localization is fundamentally dependent on the availability of a sufficient density of UE, especially near the jamming boundary. In sparsely populated areas, the system's performance will degrade significantly. The MEC algorithm may be forced to fit a circle to only a handful of widely dispersed points, resulting in a geometrically unstable solution with a large potential error [244].

To address this, the system can employ predictive modeling to augment the sparse real-world data [143]. The fusion entity could maintain a physics-informed radio propagation model of the operational area. When a jamming event is detected via a small number of sparse reports, the system could use these reports as calibration points for the model, which would then predict the most likely shape and extent of the jammed area. This approach shifts the paradigm from purely data-driven geometry to a hybrid model-based system that can produce a reasonable estimate even with limited data [147]. Furthermore, the system can implement data prioritization, assigning a higher weight or belief mass in the fusion process to reports from UE that are known to be near the estimated jamming boundary, as their data is the most valuable for refining the localization estimate.

4.4.4 System Dynamics and UE Mobility

The mobility of UE introduces a significant challenge related to data freshness. A report from a fast-moving vehicle contains location information that becomes stale within seconds. Using outdated location data in the fusion and localization algorithms can blur the perceived jamming boundary, leading to an inaccurate Jammed Area Map.

To counteract this, the fusion algorithm must be designed to be time-aware. A straightforward approach is time-weighting, where the belief mass or influence of a report decays exponentially as a function of its age. A more formal and powerful approach is to explicitly manage data freshness using the AoI metric [215, 148, 43]. AoI measures the time elapsed since the generation of the data currently being used. The system could be architected to optimize not just for detection accuracy but also to minimize the average AoI of the reports used to construct the global view. This might involve the fusion entity dynamically requesting more frequent updates from UE that are moving quickly or are in critically important locations, creating a closed-loop control system for data freshness.

The mitigations for these challenges are not isolated but form an interconnected system of

adaptive controls. For instance, the need for fresh data from a mobile UE (Challenge 4.4.4) suggests more frequent reporting, which in turn increases communication overhead (Challenge 4.4.1). This implies that the reporting rate cannot be static, which must be adaptive, increasing with UE velocity. However, a compromised UE in an FDI attack (Challenge 4.4.2) might report with anomalously high frequency. A trust system (mitigation for Challenge 4.4.2) could detect this behavior and instruct the malicious UE to lower its reporting rate, thereby mitigating both the security threat and the communication overhead it generates. This points to the necessity of a sophisticated control plane where the fusion entity can dynamically adjust the reporting parameters of individual UE based on a holistic assessment of network state, UE mobility, and UE trustworthiness.

4.4.5 Handling Multiple Jamming Sources

The proposed framework, particularly the MEC localization algorithm, is primarily designed for a single-jammer scenario. If two or more distinct jammers are active simultaneously, the set of 'JAMMED' UE will likely form a non-convex, composite shape (e.g., resembling the union of two circles). Applying the standard MEC algorithm to this entire set of points would fail, producing a single, large covering circle with a center located somewhere between the true jammer positions, providing a misleading and useless location estimate.

To handle this advanced scenario, the framework requires a preprocessing step before localization. A promising approach is to employ a spatial clustering algorithm to first partition the 'JAMMED' reports into distinct groups [214]. An algorithm like Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [68] is particularly well-suited for this task. DBSCAN can identify clusters of arbitrary shape in an unsupervised manner, making it ideal for detecting an unknown number of jammers. The workflow would be modified as follows: first, the fusion entity would apply DBSCAN to the coordinates of all high-confidence 'JAMMED' UE. If the algorithm identifies two or more distinct clusters, the system can infer the presence of multiple jammers. Subsequently, the MEC localization algorithm would be applied independently to each cluster to estimate the location of each individual jamming source [47, 145].

4.5 Chapter Summary

This chapter has presented a novel conceptual framework for collaborative intelligence designed to achieve scalable, network-wide jamming situational awareness in 5G and IoT networks. Moving beyond the inherent spatial limitations of the single-node detector detailed in previous chapters, this framework proposes a system that leverages distributed UE as a massive, collective sensor network. The core design philosophy is predicated on minimalism and scalability, using simple binary status indicators from a vast number of devices to reconstruct a high-fidelity, system-level view of the physical-layer threat landscape.

The proposed framework is built on the core principles of minimalism, scalability, and decentralized sensing, and it leverages a lightweight architecture that is well-suited to the constraints of real-world 5G and IoT deployments. The use of minimalist binary reporting, while involving a trade-off in terms of information loss at the individual UE level, is a key enabler of the framework's scalability and is more than compensated for by the richness of the data that emerges from the collective.

The core mechanisms of the framework, a robust fusion algorithm based on DST and a computationally efficient localization algorithm based on the MEC, are well-suited to the specific challenges of aggregating unreliable binary data and localizing jammers in a distributed sensing environment. The framework's ability to derive fine-grained insights into the jammer's characteristics and to handle a variety of attack scenarios has been demonstrated through a series of qualitative walkthroughs.

While a quantitative evaluation remains a subject for future work, this chapter has established a robust, well-reasoned, and qualitatively analyzed blueprint for a next-generation collaborative security system. A comprehensive analysis of the significant operational challenges, including communication overhead, data reliability, UE density, system dynamics, and the presence of multiple jamming sources, has also been conducted. These challenges, along with the proposed mitigation strategies, are not presented as insurmountable flaws but as critical areas for the future research and engineering required to translate this conceptual framework into a practical, deployed system. Ultimately, this lightweight, collaborative approach represents a highly practical and effective path towards bolstering the security and resilience of future wireless networks against the persistent threat of jamming attacks.

Discussion and Conclusion

The preceding chapters have detailed a multi-stage research effort to address the challenge of jamming in 5G wireless systems, progressing from foundational on-device detection to a scalable, collaborative intelligence framework. This chapter synthesizes these contributions, reflects on their practical implications and inherent limitations, and outlines promising directions for future work. It aims to place the specific findings of this thesis within the broader context of wireless security and ML research, before offering final concluding remarks.

5.1 Synthesis of Contributions

This thesis was motivated by a critical reality gap identified in the existing literature on ML-based wireless security. Prior research has been characterized by an over-reliance on sanitized, simulated data, the use of simplistic threat models, and the prevalence of centralized processing architectures that are ill-suited for real-world deployment. The central objective of this work, therefore, has been to develop and rigorously validate a scalable, robust, and practical framework for detecting advanced physical layer threats within the complex and dynamic environments of 5G.

The contributions presented in the preceding chapters are not isolated studies but form a single, coherent, and synergistic solution arc that addresses the problem of 5G security with progressively increasing levels of sophistication and scale. The research journey begins in Chapter 2 by establishing the foundational feasibility of performing sophisticated, hardware-agnostic jammer detection directly on COTS devices using natively available system logs.

Recognizing that the performance of any such on-device model is fundamentally constrained by data availability and the non-stationary nature of the wireless medium, the work in Chapter 3 confronts the critical challenge of model generalization. It introduces a novel generative data augmentation strategy designed to enhance detector robustness against the inevitable distribution shifts encountered in operational environments.

Finally, acknowledging the inherent spatial limitations of any single sensor, Chapter 4 proposes a conceptual framework to scale the solution from an individual device to network-wide, collaborative intelligence, thereby transforming isolated, localized detections into actionable, system-level situational awareness. This three-stage progression, from on-device feasibility through data-driven robustness to collaborative scaling, represents a unified contribution. It provides a complete, end-to-end blueprint for a new generation of wireless security systems that are decentralized, data-resilient, and scalable by design.

5.2 Interpretation of Key Findings

This section moves beyond a mere summary of the results presented in preceding Chapters to provide a deeper interpretation of their scientific and practical significance. Each key finding is analyzed in the context of the thesis’s overarching goals, articulating what these results truly signify for the field of wireless security.

5.2.1 The Efficacy of On-Device, Log-Based Detection

The empirical evaluation detailed in Chapter 2 demonstrated that the SHIELD framework, when paired with lightweight ML classifiers, could achieve a detection accuracy of approximately 94% on an unseen robustness test set. This performance was achieved against a sophisticated, power-modulated jammer (a threat model deliberately designed to evade naive detection methods) and stands in stark contrast to the catastrophic failure of benchmarked methods from prior literature under the same conditions.

The significance of this finding extends far beyond the numerical accuracy. It serves as a fundamental validation that native system logs, generated by default on COTS devices, are a rich, yet largely untapped, source of high-fidelity security telemetry. The success of the SHIELD pipeline reveals that the information required to detect advanced physical layer attacks is not exclusively confined to the raw radio frequency domain, accessible only via specialized and costly hardware like spectrum analyzers. Rather, this information is already present and latent within the operational metrics logged by the device’s own modem and operating system. It simply requires the appropriate feature engineering and ML techniques to be unlocked.

This finding represents a potential paradigm shift, moving the problem of physical layer security from the domain of expensive, specialized hardware to the domain of scalable, deployable software. A deeper analysis of the results reveals a crucial causal relationship that underpins this success. The failure of the benchmarked methods was not a result of a poor choice of classifier. Indeed, they often used the same algorithms, such as SVM or RnF. It was a direct consequence of the poverty of their input features, which typically consisted of raw, instantaneous metrics like PDR or RSSI. The power-modulated jammer was specifically engineered to defeat such simple monitoring. Its fluctuating power profile means that at any given moment, its signal strength might fall within a range considered normal, rendering static thresholding ineffective.

The success of the SHIELD pipeline was therefore not merely in the application of a ML model, but in its sophisticated preprocessing and feature engineering methodology. The use of a sliding-window aggregation stack to compute higher-order statistical features, including the standard deviation and kurtosis of signal metrics over time, was the critical step. These features do not measure the signal's level, but its volatility and the presence of extreme outliers. This transformed the data into a new feature space where the jammer's signature, previously hidden, became mathematically apparent and linearly separable for the classifiers. The contribution, therefore, lies as much in the intelligent data transformation as in the final classification model itself.

5.2.2 The Power of the Synthetic-Only Training Strategy

The cross-scenario evaluation in Chapter 3 produced a remarkable and counter-intuitive result: an anomaly detector trained exclusively on synthetic time-series data generated by the GANSec framework achieved an accuracy of 92.1% on an unseen test environment. This dramatically outperformed the model trained on the original, real data from the source domain, which collapsed to a mere 78.1% accuracy when faced with the same distribution shift.

This finding has profound implications for how robust ML systems are developed. It demonstrates that under conditions of data scarcity and environmental dynamism, a high-fidelity synthetic dataset can be more valuable for training than a limited real-world dataset. The GAN, by learning the underlying statistical distribution of the real data, was able to generate a synthetic dataset that was not only realistic but also more diverse and comprehensive than the original. This richer dataset effectively exposed the downstream detector to a wider range of plausible variations than were present in the limited real data, vaccinating it against the statistical shifts it would later encounter in the new, unseen test environment. The synthetic data did not just replicate the training environment, but also provided a more complete map of the problem space to enable superior generalization.

This success fundamentally reframes the data acquisition bottleneck, a core challenge identified in the literature review and detailed in Chapter 3. The conventional approach to building robust models is predicated on the collection of massive datasets, a process that is often physically difficult, economically prohibitive, and legally complex in the security domain. The success of the synthetic-only strategy proposes a powerful alternative paradigm. Instead of pursuing massive-scale real-world data collection, an organization can focus on acquiring a small, high-quality "seed" dataset. This seed data is then used to train a tailored, high-fidelity generative model like GANSec, which in turn can produce a virtually unlimited, diverse, and perfectly labeled synthetic dataset for training robust downstream classifiers. This approach decouples the development of resilient security models from the need for big data collection, shifting the focus from data acquisition to smart data generation. This has the potential to accelerate research and development in any domain where data is scarce, sensitive, or difficult to obtain.

5.2.3 The Viability of Lightweight Collaborative Intelligence

Chapter 4 presented a conceptual framework for a collaborative intelligence system capable of generating network-wide situational awareness, including jammer localization, from minimalist binary inputs from a distributed population of UE. The conceptual soundness of this design signifies a viable path toward creating truly scalable, decentralized security systems for the massive IoT networks of the future.

The core insight of this framework lies in a deliberate and strategic architectural trade-off: it sacrifices information fidelity at the level of the individual node to achieve massive scalability and minimize the operational burden on resource-constrained devices. This design is an embodiment of the principle of emergent properties, where the aggregation of a vast number of simple, low-fidelity, and potentially unreliable data points can be fused to create a complex, high-fidelity, and reliable system-level understanding. While a single binary report is of little value, the spatiotemporal patterns formed by thousands or millions of such reports can reveal the scope, location, and behavior of a physical layer threat with high confidence.

This framework provides a blueprint for what is effectively a crowdsourced physical layer sensor network. Traditional security monitoring relies on the deployment of a limited number of dedicated, expensive probes by the network operator. The proposed collaborative framework, in contrast, leverages the sensing capabilities of devices that are already deployed throughout the network for other purposes, such as smartphones, smart home devices, and industrial sensors. By enabling these devices to contribute a single, low-cost bit of information, the framework creates a distributed security instrument of unprecedented scale and density at virtually no additional hardware cost. This approach democratizes the sensing process, providing a level of spatial granularity and real-time coverage that would be economically and logistically impossible to achieve with dedicated, operator-owned hardware. This has significant implications not only for security monitoring but also for related applications such as dynamic network performance monitoring and radio environment mapping.

5.3 Implications of the Research

The findings of this thesis have broader implications that extend beyond the immediate research questions, impacting practical applications, research methodologies, and theoretical understanding in the fields of network security and ML.

5.3.1 Practical Implications

The presented work provides a direct and practical pathway for network operators, device manufacturers, and security professionals to develop and deploy a new class of low-cost, scalable, and purely software-based security monitoring tools. The on-device detector developed in Chapter 2 could be integrated into base station software stacks or as part of the firmware on UE and IoT

devices, providing a first line of defense with no additional hardware cost. The collaborative intelligence framework from Chapter 4 could be implemented by mobile network operators as a network-level security service, offering them unprecedented, real-time visibility into the physical layer threat landscape across their entire infrastructure. This could transform security from a reactive, incident-based process into a proactive, data-driven operational function.

While the SHIELD and GANSec frameworks detailed in this thesis employed an off-line training strategy, the principles have been extended to demonstrate real-world applicability. A demonstration project of online jamming detection was implemented, which successfully extracts real-time Android logs, processes them into feature buffers, and predicts the presence of a jammer using a pre-trained LSTM model. The pipeline for this real-time implementation, as depicted in Figure 5.1, demonstrates a practical workflow from log ingestion via ADB to feature processing and final classification. This proof-of-concept, available publicly at the DEMO_Android_JD repository, validates the feasibility of deploying these log-based detection techniques for live, on-device monitoring, bridging the gap between academic research and practical, real-time security applications.

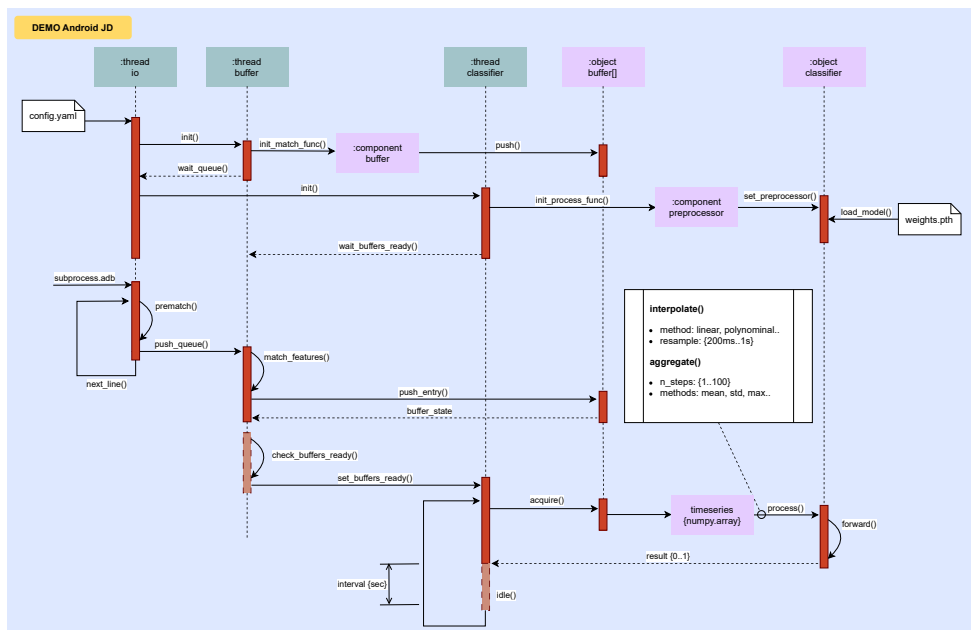


Figure 5.1: Sequence diagram of the real-time jamming detection pipeline prototype, implemented on Android logs.

This research is also a direct contribution to the EU-funded project MLSysOps, specifically contributing to the wireless network anomaly detection module for the project's deliverables. As part of this project, we are collaborating with industry partner NTT Data to explore the practical applicability of our jamming detection approach on industrial-grade hardware, such as

the Teltonika RUTX50 5G router. The ongoing work and code for this industrial use case are available at https://github.com/jxu96/teltonika_usecase.

5.3.2 Methodological Implications

The two-dataset evaluation strategy, which was designed and employed in Chapter 2 to rigorously test model generalization, should be considered a standard methodology for robustness testing in wireless security research. By using a separate robustness test set collected under demonstrably different environmental conditions, this approach provides a far more stringent and realistic assessment of a model’s ability to generalize than a simple train-test split on a single, homogeneous dataset.

Furthermore, this work establishes a valuable methodological blueprint for processing and feature engineering from native, real-world system logs. The multi-stage pipeline, encompassing synchronized cross-end data collection, robust interpolation for asynchronous data, and sliding-window aggregation for temporal feature extraction, provides a replicable process for transforming noisy, unstructured log files into a high-dimensional feature space suitable for ML. This stands in contrast to many prior studies that rely on clean, pre-structured data from simulators, offering a more realistic methodology for researchers working with messy, operational data. Adopting these methodologies more broadly would help to combat the reality gap that plagues the field, and to encourage the development of more resilient and genuinely deployable systems for reproducible research.

5.3.3 Theoretical Implications

This work contributes to several broader theoretical discussions. The success of the GANSec framework in Chapter 3, particularly the task-oriented (CO-GAN) configuration, shows evident effectiveness of task-oriented generative models. It provides strong empirical evidence for the theory that under conditions of data scarcity and distribution shift, a well-trained generative model can produce a synthetic dataset that is more valuable for training a robust classifier than the limited real-world data it was trained on. This challenges the conventional big data paradigm, contributing to an emerging theory of smart data where the focus shifts from massive data collection to the intelligent synthesis of diverse and representative training examples.

Furthermore, the collaborative framework in Chapter 4 contributes to the theory of decentralized intelligence. It provides a strong conceptual model for the theory of emergent properties in large-scale distributed systems. It demonstrates how reliable, high-level intelligence, such as threat localization, can emerge from the aggregation of a massive number of simple, unreliable, and low-fidelity binary sensor inputs. This empirically grounds the principle that in such systems, the quantity and spatial diversity of sensors can effectively compensate for, and ultimately overcome, the quality and information limitations of individual sensing units, offering a new perspective on designing resilient systems from unreliable components.

5.4 Limitations of the Presented Work

A consolidated and self-critical assessment of the research's boundaries is essential for contextualizing the contributions and guiding future work. The limitations of this thesis can be organized thematically.

5.4.1 Scope of Experimental Validation

The empirical work presented in Chapters 2 and 3, while utilizing real hardware, was conducted within a controlled laboratory testbed. This experimental environment was necessarily limited in scope, involving a single model of a COTS smartphone, a single software implementation of a gNB (srsRAN [225]), and one specific, yet sophisticated power-modulated jammer. It is important to clarify that the intent of this research was not to propose a universal ML model capable of covering all possible attack scenarios or predicting entirely new classes of attacks. Rather, the goal was to prove the effectiveness of the proposed methodology. This limitation is therefore reasonable, by conducting a rigorous evaluation with representative hardware and a challenging threat model, we have demonstrated the viability of the on-device, log-based detection and generative augmentation approach. This foundational proof of concept can, and should, be extended to more diverse scenarios in future work, e.g., a wider range of devices from different manufacturers, diverse network vendor equipment, and a broader array of jamming attack strategies.

5.4.2 Assumptions in the Generative and Collaborative Models

The models developed in this thesis operate on a set of underlying assumptions. A primary limitation of the GANSec model is that it serves as a data augmentation tool, not a complete replacement for physical data collection. The collection of a small-amount, high-quality, and representative seed dataset remains a necessary prerequisite, as the model's ability to generate realistic data is fundamentally constrained by the quality of this initial input. Furthermore, the model can only learn to generate variations of patterns present in the seed data, it cannot predict or generate patterns for entirely new classes of attacks it has not been exposed to. Secondly, the generation of synthetic data, as with most GAN-based approaches, presents some randomness and a notable lack of interpretability. The process is a black box that is difficult to explain why a particular sample was generated or to guarantee that specific feature interactions are captured. Further regulation methods could be investigated to improve the transparency of the generation process. Third, the current condition vector c used in GANSec's embedding layer is limited to higher-level contextual features like device model and location. The inclusion of more granular features from the physical layer conditions has the potential to guide the generator in revealing more realistic and nuanced data patterns.

For the collaborative framework, the conceptual analysis assumes a baseline of honest

participation from the UE. While resilience to FDI attacks is discussed as a critical area for future work, the current conceptual model does not include fully implemented mechanisms to defend against coordinated, malicious reporting from compromised devices.

5.4.3 Lack of Quantitative Evaluation for the Collaborative Framework

The framework for collaborative intelligence in Chapter 4 is presented as a conceptual and architectural blueprint, and its viability is argued through qualitative analysis. Its performance has not been validated through large-scale simulation or a real-world pilot deployment. This is a significant limitation, primarily driven by two practical constraints. First, the logistical challenge and cost of acquiring and deploying the large number of physical devices required for a meaningful quantitative evaluation were prohibitive within the scope of this research. Second, a simulation-based evaluation was not a viable alternative because existing network simulators typically do not provide access to the specific, low-level, and often proprietary system log features that form the basis of the on-device binary detector. Without the ability to generate these requisite features, it is impossible to simulate the foundational input for the collaborative system.

5.5 Future Research Directions

Based on the implications and limitations identified, a clear and compelling roadmap for future research can be proposed. This roadmap consolidates the future directions suggested in each chapter into a cohesive set of research avenues.

5.5.1 Expanding Experimental Validation

The most immediate and critical next step is to validate the SHIELD and GANSec frameworks in more diverse and realistic environments. This involves a multi-pronged effort to:

- Conduct experiments with a wide variety of smartphone and IoT device models to assess the portability of the log-parsing and feature engineering pipeline.
- Test the systems against different 5G core and RAN software implementations to ensure vendor-agnostic performance. Specifically, we plan to leverage the FIT R2Lab testbed [180], which provides a versatile environment with programmable nodes suitable for emulating 5G scenarios, and introducing controlled jamming interference.
- In the mid-term, efforts could be expanded to develop and evaluate the detectors against a more comprehensive library of jamming attack types, including dynamic strategies like frequency-hopping and protocol-aware jammers [37, 29, 73] that target specific control channels.

- Our present approach primarily identifies whether a jamming attack is active, without distinguishing its intensity or specific interference patterns. More detailed detection strategies could provide insights such as jammer power levels or temporal patterns of interference. This information would be especially valuable for refined countermeasures, such as altering scheduling or power-control parameters on the gNB side.
- Ultimately, this line of research should culminate in the deployment and evaluation of the system in a live, city-scale 5G testbed to assess its performance within the complexity and unpredictability of a real operational network.

5.5.2 Advancing the ML Methodologies

The ML models used in this thesis provide a strong baseline but can be extended to create more adaptive and powerful systems. In the short-term, future work could explore:

- **Online Learning:** Modifying the on-device detectors to incorporate online or continual learning capabilities [161, 174]. This would allow the models to adapt in real-time to gradual changes in the environment or novel threat patterns, moving from a static, pre-trained model to a lifelong learning system.
- **New Generative Models:** Investigating the application of more recent and powerful generative architectures, such as Denoising Diffusion Probabilistic Models (mentioned in Section 3.1.4), to the task of wireless time-series synthesis. These models may be capable of generating synthetic data with even higher fidelity and diversity, potentially leading to further improvements in detector robustness [48, 230, 113].

5.5.3 Hardening the Collaborative Framework

The long-term goal is to move the conceptual framework from Chapter 4 into a robust and secure implementation. Key research tasks and challenges include:

- Conducting large-scale simulations to quantitatively model the system's performance, focusing on the trade-offs between UE density, reporting strategy (e.g., periodic vs. event-triggered), communication overhead, and localization accuracy.
- Developing and testing the spatial clustering algorithms (e.g., DBSCAN) required as a preprocessing step to enable the localization of multiple, simultaneous jamming sources.
- Implementing and evaluating the proposed FDI resilience mechanisms, such as lightweight cryptographic signatures for report authentication and the development of a dynamic trust and reputation system to down-weight suspicious nodes.

5.5.4 Moving from Detection to Autonomous Mitigation

The final goal of this research is to close the loop from passive detection to active, autonomous response. Future work should investigate how the high-level, network-wide intelligence generated by the collaborative framework can be used as a trigger for automated mitigation actions [219]. This could involve developing control-plane logic that, upon confident detection and localization of a jammer, autonomously initiates countermeasures such as instructing affected UE to perform localized frequency hopping [41, 136], dynamically adjusting transmit power levels [150], or re-routing critical traffic through unaffected areas of the network [190].

5.6 Concluding Remarks

This thesis has confronted the critical and evolving challenge of securing next-generation wireless networks against physical layer attacks. It has systematically addressed and dismantled the barriers that have historically limited practical and scalable detection systems by proposing and validating a novel, three-stage framework that progresses from on-device feasibility to data-driven robustness and finally to collaborative, network-wide intelligence.

When evaluated from a short-term perspective, this work serves as a foundational exploration into the use of ML-based self-intelligence for on-device anomaly sensing. It initiates a practical line of research into AI for security that moves beyond simulation, providing tangible methodologies for real-world data collection, generative augmentation, and robust model validation. This immediate contribution is already being realized within the EU-funded MLSysOps project, where the developed techniques form the basis of the wireless network anomaly detection module.

From a long-term perspective, this thesis promises more than an incremental improvement. It lays the architectural groundwork for a future of large-scale, intelligent, and collaboratively anomaly-resilient connected networks. The conceptual framework for minimalist, decentralized intelligence provides a scalable blueprint for transforming billions of isolated endpoints into a collective, self-healing sensor network. By grounding this vision in a practical MLOps context through collaborations with industrial partners, this work not only advances the state-of-the-art but also helps forge a viable path toward ensuring the availability and integrity of the critical 5G and IoT infrastructure upon which our future increasingly depends.

Bibliography

- [1] 3GPP. 5G; NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (3GPP TS 38.101-1 version 18.10.0 Release 18). Technical Specification (TS).
- [2] 3GPP. 5G; NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone (3GPP TS 38.101-2 version 18.10.0 Release 18). Technical Specification (TS).
- [3] 3GPP. 5G; Security architecture and procedures for 5G System (3GPP TS 33.501 version 17.12.0 Release 17). Technical Specification (TS). Jan. 2024. (Visited on 04/11/2024).
- [4] 3GPP. 5G; Security aspects of enhancement of support for enabling edge applications (3GPP TS 33.558 version 17.3.0 Release 17). Technical Specification (TS). Jan. 2023.
- [5] 3GPP. 5G; System architecture for the 5G System (5GS) (3GPP TS 23.501 version 17.11.0 Release 17). Technical Specification (TS). Jan. 2024.
- [6] 3GPP. Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Numbering, addressing and identification (3GPP TS 23.003 version 17.10.0 Release 17). Technical Specification (TS). July 2023.
- [7] Muhammad Aaqib et al. "IoT trust and reputation: a survey and taxonomy". In: *Journal of Cloud Computing* 12.1 (2023), p. 42.
- [8] Mahmoud Abbasi et al. "Security in the Internet of Things Application Layer: Requirements, Threats, and Solutions". In: *IEEE Access* 10 (2022), pp. 97197–97216. doi: 10.1109/ACCESS.2022.3205351.
- [9] Garima Agrawal, Amardeep Kaur, and Sowmya Myneni. "A review of generative models in generating synthetic attack data for cybersecurity". In: *Electronics* 13.2 (2024), p. 322.
- [10] Rami Ahmad, Raniyah Wazirali, and Tarik Abu-Ain. "Machine learning for wireless sensor networks security: An overview of challenges and issues". In: *Sensors* 22.13 (2022), p. 4730.
- [11] Adnan Ahmed et al. "Stealth jamming attack in WSNs: Effects and countermeasure". In: *IEEE Sensors Journal* 18.17 (2018), pp. 7106–7113.
- [12] Hafiza Anisa Ahmed, Anum Hameed, and Narmeen Zakaria Bawany. "Network intrusion detection using oversampling technique and machine learning algorithms". In: *PeerJ Computer Science* 8 (2022), e820.
- [13] NAR Akinyele et al. "Leveraging generative Artificial Intelligence (AI) for cybersecurity: Analyzing diffusion models in detecting and mitigating cyber threats". In: *GSC Advanced Research and Reviews* 21.2 (2024), pp. 001–014.
- [14] Somya Abdulkarim Alhandi, Hazalila Kamaludin, and Nayef Abdulwahab Mohammed Alduais. "Trust Evaluation Model in IoT Environment: A Comprehensive Survey". In: *IEEE Access* 11 (2023), pp. 11165–11182. doi: 10.1109/ACCESS.2023.3240990.

- [15] Iman Almomani, Bassam Al-Kasasbeh, and Mousa AL-Akhras. "WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks". In: *Journal of Sensors* 2016.1 (2016), p. 4731953. doi: <https://doi.org/10.1155/2016/4731953>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1155/2016/4731953>.
- [16] Mamyr Altaibek et al. "A Survey of Cross-Layer Security for Resource-Constrained IoT Devices". In: *Applied Sciences* 15.17 (2025). issn: 2076-3417. doi: 10.3390/app15179691. url: <https://www.mdpi.com/2076-3417/15/17/9691>.
- [17] Dure Adan Ammara, Jianguo Ding, and Kurt Tutschku. "Synthetic Network Traffic Data Generation: A Comparative Study". In: *arXiv preprint arXiv:2410.16326* (2024).
- [18] Kenneth Li-Minn Ang, Jasmine Kah Phooi Seng, and Adamu Murtala Zungeru. "Optimizing Energy Consumption for Big Data Collection in Large-Scale Wireless Sensor Networks With Mobile Collectors". In: *IEEE Systems Journal* 12.1 (2018), pp. 616–626. doi: 10.1109/JSYST.2016.2630691.
- [19] Youness Arjoune and Saleh Faruque. "Real-time machine learning based on hoeffding decision trees for jamming detection in 5G new radio". In: *2020 IEEE International Conference on Big Data (Big Data)*. IEEE. 2020, pp. 4988–4997.
- [20] Youness Arjoune and Saleh Faruque. "Smart jamming attacks in 5G new radio: A review". In: *2020 10th annual computing and communication workshop and conference (CCWC)*. IEEE. 2020, pp. 1010–1015.
- [21] Youness Arjoune et al. "A Novel Jamming Attacks Detection Approach Based on Machine Learning for Wireless Communication". In: *2020 International Conference on Information Networking (ICOIN)*. 2020, pp. 459–464. doi: 10.1109/ICOIN48656.2020.9016462.
- [22] Mehmet Arslan et al. "SMOTE and gaussian noise based sensor data augmentation". In: *2019 4th international conference on computer science and engineering (UBMK)*. IEEE. 2019, pp. 1–5.
- [23] Orchu Aruna et al. "Multi-Level Node Behavior Pattern Analysis for Malicious Node Detection with False Alarm Reduction in Wireless Sensor Networks." In: *International Journal of Safety & Security Engineering* 15.6 (2025).
- [24] Sriranga Narasimha Gandhi Aryavalli and Hemantha Kumar. "Top 12 layer-wise security challenges and a secure architectural solution for Internet of Things". In: *Computers and Electrical Engineering* 105 (2023), p. 108487. issn: 0045-7906. doi: <https://doi.org/10.1016/j.compeleceng.2022.108487>. url: <https://www.sciencedirect.com/science/article/pii/S0045790622007029>.
- [25] Ghazal Asemian et al. "Over-The-Air Double-Threshold Deep Learner for Jamming Detection in 5G RF domain". In: *arXiv preprint arXiv:2403.02645* (2024).
- [26] Roshan Ayyalasomayajula et al. "Deep learning based wireless localization for indoor navigation". In: *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 2020, pp. 1–14.
- [27] Nur Athirah Azhar et al. "An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis". In: *IEEE Transactions on Knowledge and Data Engineering* 35.7 (2022), pp. 6651–6672.
- [28] Sachi D Babar, Neeli R Prasad, and Ramjee Prasad. "Jamming attack: Behavioral modelling and analysis". In: *Wireless VITAE 2013*. IEEE. 2013, pp. 1–5.
- [29] Sanjoy Basak et al. "Spectrum prediction for protocol-aware RF jamming". In: *IEEE Transactions on Cognitive Communications and Networking* 10.2 (2023), pp. 363–373.

- [30] Md Abul Bashar and Richi Nayak. "TAnoGAN: Time Series Anomaly Detection with Generative Adversarial Networks". In: *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2020, pp. 1778–1785. doi: 10.1109/SSCI47803.2020.9308512.
- [31] Narmeen Zakaria Bawany, Jawwad A Shamsi, and Khaled Salah. "DDoS attack detection and mitigation using SDN: methods, practices, and solutions". In: *Arabian Journal for Science and Engineering* 42.2 (2017), pp. 425–441.
- [32] Aborisade IYANUOLUWA Blessing, Adam Chole, and Yusuf Themmah. "Signature-Based vs. Anomaly-Based Detection". In: (2023).
- [33] Carlo Alberto Boano, Marco Cattani, and Kay Römer. "Impact of temperature variations on the reliability of LoRa". In: *Proc. 7th Int. Conf. Sensor Netw.* 2018, pp. 39–50.
- [34] Andrey Bogdanov et al. "PRESENT: An ultra-lightweight block cipher". In: *International workshop on cryptographic hardware and embedded systems*. Springer. 2007, pp. 450–466.
- [35] Chaimaa Boudagdigue et al. "Trust Management in Industrial Internet of Things". In: *IEEE Transactions on Information Forensics and Security* 15 (2020), pp. 3667–3682. issn: 1556-6021. (Visited on 04/17/2024).
- [36] Badre Bousalem et al. "Deep learning-based smart radio jamming attacks detection on 5G V2I/V2N communications". In: *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE. 2023, pp. 7139–7144.
- [37] Emilie Bout et al. "FOLPETTI: A novel multi-armed bandit smart attack for wireless networks". In: *Proceedings of the 17th International Conference on Availability, Reliability and Security*. 2022, pp. 1–10.
- [38] Hamda Bouzabia, Tri Nhu Do, and Georges Kaddoum. "Deep learning-enabled deceptive jammer detection for low probability of intercept communications". In: *IEEE Systems Journal* 17.2 (2022), pp. 2166–2177.
- [39] Daniele Canavese et al. "Security at the edge for resource-limited IoT devices". In: *Sensors* 24.2 (2024), p. 590.
- [40] Hanqun Cao et al. "A survey on generative diffusion models". In: *IEEE transactions on knowledge and data engineering* 36.7 (2024), pp. 2814–2830.
- [41] Chih-Min Chao and Wei-Che Lee. "Load-aware anti-jamming channel hopping design for cognitive radio networks". In: *Computer Networks* 184 (2021), p. 107681.
- [42] N. V. Chawla et al. "SMOTE: Synthetic Minority Over-sampling Technique". In: *Journal of Artificial Intelligence Research* 16 (June 2002), pp. 321–357. issn: 1076-9757. doi: 10.1613/jair.953. url: <http://dx.doi.org/10.1613/jair.953>.
- [43] Shengzhan Chen et al. "Data Freshness Quality in Privacy-Enabled Mobile Crowdsensing: Design Aspects and Challenges". In: *Electronics* 14.7 (2025), p. 1365.
- [44] Zhenguo Chen, Liqin Tian, and Chuang Lin. "Trust model of wireless sensor networks and its application in data fusion". In: *Sensors* 17.4 (2017), p. 703.
- [45] Long Cheng et al. "A survey of localization in wireless sensor network". In: *International Journal of Distributed Sensor Networks* 8.12 (2012), p. 962523.
- [46] Tianzhen Cheng, Ping Li, and Sencun Zhu. "An algorithm for jammer localization in wireless sensor networks". In: *2012 IEEE 26th international conference on advanced information networking and applications*. IEEE. 2012, pp. 724–731.

- [47] Tianzhen Cheng, Ping Li, and Sencun Zhu. "Multi-jammer localization in wireless sensor networks". In: *2011 Seventh International Conference on Computational Intelligence and Security*. IEEE. 2011, pp. 736–740.
- [48] Guoxuan Chi et al. "RF-Diffusion: Radio Signal Generation via Time-Frequency Diffusion". In: *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. ACM MobiCom '24. Association for Computing Machinery, 2024. doi: 10.1145/3636534.3649348.
- [49] Merlin Chlosta et al. "LTE security disabled: misconfiguration in commercial networks". In: *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*. WiSec '19. ACM, May 2019, pp. 261–266. ISBN: 978-1-4503-6726-4. (Visited on 04/12/2024).
- [50] Kostas Chounos, Katerina Kyriakou, and Thanasis Korakis. "Scalability and Performance Evaluation of IEEE 802.11Ah IoT Deployments: A Testbed Approach". In: *2025 21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2025, pp. 858–865. doi: 10.1109/DCOSS-IoT65416.2025.00131.
- [51] T Charles Clancy. "Efficient OFDM denial: Pilot jamming and pilot nulling". In: *2011 IEEE International Conference on Communications (ICC)*. IEEE. 2011, pp. 1–5.
- [52] Mauro Conti et al. "Internet of Things security and forensics: Challenges and opportunities". In: *Future Generation Computer Systems* 78 (Jan. 2018), pp. 544–546. ISSN: 0167-739X. (Visited on 04/12/2024).
- [53] Juan David Arias Correa, Alex Sandro Roschildt Pinto, and Carlos Montez. "Lossy data compression for iot sensors: A review". In: *Internet of Things* 19 (2022), p. 100516.
- [54] Alejandro Cortés-Leal et al. "Performance metric analysis for a jamming detection mechanism under collaborative and cooperative schemes in industrial wireless sensor networks". In: *Sensors* 22.1 (2021), p. 178.
- [55] Damien Dablain, Bartosz Krawczyk, and Nitesh V. Chawla. "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data". In: *IEEE Transactions on Neural Networks and Learning Systems* 34.9 (2023), pp. 6390–6404. doi: 10.1109/TNNLS.2021.3136503.
- [56] Adrian Dabrowski et al. "IMSI-catch me if you can: IMSI-catcher-catchers". In: *Proceedings of the 30th annual computer security applications Conference*. 2014, pp. 246–255.
- [57] Muataz Salam Al-Daweri et al. "An analysis of the KDD99 and UNSW-NB15 datasets for the intrusion detection system". In: *Symmetry* 12.10 (2020), p. 1666.
- [58] Carolina Del-Valle-Soto et al. "A low-cost jamming detection approach using performance metrics in cluster-based wireless sensor networks". In: *Sensors* 21.4 (2021), p. 1179.
- [59] Carolina Del-Valle-Soto et al. "Adaptive Jamming Mitigation for Clustered Energy-Efficient LoRa-BLE Hybrid Wireless Sensor Networks". In: *Sensors* 25.6 (2025), p. 1931.
- [60] Carolina Del-Valle-Soto et al. "New detection paradigms to improve wireless sensor network performance under jamming attacks". In: *Sensors* 19.11 (2019), p. 2489.
- [61] Asside Christian Djedouboum et al. "Big data collection in large-scale wireless sensor networks". In: *Sensors* 18.12 (2018), p. 4474.
- [62] Peng Du and George Roussos. "Adaptive communication techniques for the internet of things". In: *Journal of Sensor and Actuator Networks* 2.1 (2013), pp. 122–155.

- [63] Dr. Lachit Dutta and Swapna Bharali. "TinyML Meets IoT: A Comprehensive Survey". In: *Internet of Things* 16 (2021), p. 100461. issn: 2542-6605. doi: <https://doi.org/10.1016/j.iot.2021.100461>. URL: <https://www.sciencedirect.com/science/article/pii/S2542660521001025>.
- [64] Hrishikesh Dutta, Amit Kumar Bhuyan, and Subir Biswas. "Using Multi-Armed Bandit Learning for Thwarting MAC Layer Attacks in Wireless Networks". In: *IEEE/ACM Transactions on Networking* (2024).
- [65] Peter Eigenschink et al. "Deep generative models for synthetic data: A survey". In: *IEEE Access* 11 (2023), pp. 47304–47320.
- [66] Khaled El Emam, Lucy Mosquera, and Richard Hoptroff. *Practical synthetic data generation: balancing privacy and the broad availability of data*. O'Reilly Media, 2020.
- [67] Tugba Erpek, Yalin E Sagduyu, and Yi Shi. "Deep learning for launching and mitigating wireless jamming attacks". In: *IEEE Transactions on Cognitive Communications and Networking* 5.1 (2018), pp. 2–14.
- [68] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *kdd*. Vol. 96. 34. 1996, pp. 226–231.
- [69] Jianhua Fan et al. "Identification and localization of the jammer in wireless sensor networks". In: *The Computer Journal* 62.10 (2019), pp. 1515–1527.
- [70] Zongwen Fan et al. "Sampling-based machine learning models for intrusion detection in imbalanced dataset". In: *Electronics* 13.10 (2024), p. 1878.
- [71] Zhutian Feng and Cunqing Hua. "Machine Learning-based RF Jamming Detection in Wireless Networks". In: *2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC)*. 2018, pp. 1–6. doi: 10.1109/SSIC.2018.8556709.
- [72] Carmen Fernandez-Gago, Francisco Moyano, and Javier Lopez. "Modelling trust dynamics in the Internet of Things". In: *Information Sciences* 396 (2017), pp. 72–82.
- [73] Maya E Flores et al. "Implementation and Evaluation of a Smart Uplink Jamming Attack in a Public 5G Network". In: *IEEE Access* (2023).
- [74] Henrik Forssell et al. "UE-Assisted Jamming Detection in 5G NR". In: *ICC 2023 - IEEE International Conference on Communications*. 2023, pp. 5216–5220. doi: 10.1109/ICC45041.2023.10279513.
- [75] Jerome G Gacu et al. "Application of Artificial Intelligence in Hydrological Modeling for Streamflow Prediction in Ungauged Watersheds: A Review". In: *Water* 17.18 (2025), p. 2722.
- [76] R Ganesamoorthy et al. "The Ethical Implications and Societal Impact of Widespread WSN Deployments". In: *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*. IEEE. 2024, pp. 1–6.
- [77] Zijun Gao, Haibao Liu, and Lingbo Li. "Data augmentation for time-series classification: An extensive empirical study and comprehensive survey". In: *Journal of Artificial Intelligence Research* 83 (2025).
- [78] Gianni Giorgetti, Sandeep Kumar S Gupta, and Gianfranco Manes. "Understanding the limits of RF-based collaborative localization". In: *IEEE/ACM Transactions on Networking* 19.6 (2011), pp. 1638–1651.

- [79] Patrik Goldschmidt and Daniela Chudá. "Network intrusion datasets: a survey, limitations, and recommendations". In: *Computers & Security* (2025), p. 104510.
- [80] Ian J Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).
- [81] Mahesh Kumar Goyal and R Chaturvedi. "Synthetic Data Revolutionizes Rare Disease Research: How Large Language Models and Generative AI are Overcoming Data Scarcity and Privacy Challenges". In: *International Journal on Recent and Innovation Trends in Computing and Communication* 11.11 (2023), pp. 1368–1380.
- [82] Mandeep Goyal and Qusay H Mahmoud. "A systematic review of synthetic data generation techniques using generative AI". In: *Electronics* 13.17 (2024), p. 3509.
- [83] Kanika Grover, Alvin Lim, and Qing Yang. "Jamming and anti-jamming techniques in wireless networks: a survey". In: *International Journal of Ad Hoc and Ubiquitous Computing* 17.4 (2014), pp. 197–215.
- [84] Jorge Luis Guerra, Carlos Catania, and Eduardo Veas. "Datasets are not enough: Challenges in labeling network traffic". In: *Computers & Security* 120 (2022), p. 102810.
- [85] Omer Melih Gul et al. "Secure Industrial IoT Systems via RF Fingerprinting Under Impaired Channels With Interference and Noise". In: *IEEE Access* 11 (2023), pp. 26289–26307. DOI: 10.1109/ACCESS.2023.3257266.
- [86] Maanak Gupta et al. "From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy". In: *IEEE access* 11 (2023), pp. 80218–80245.
- [87] Tien Ngoc Ha and Daniel Romero. "Spoofing Detection in the Physical Layer with Graph Neural Networks". In: *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*. IEEE. 2024, pp. 1–6.
- [88] Marouane Hachimi et al. "Multi-stage jamming attacks detection using deep learning combined with kernelized support vector machine in 5G cloud radio access networks". In: *2020 international symposium on networks, computers and communications (ISNCC)*. IEEE. 2020, pp. 1–5.
- [89] Jacob Haislip et al. "The economic cost of cybersecurity breaches: A broad-based analysis". In: *Workshop on the economics of information security (WEIS)*. Vol. 1. 2019, p. 37.
- [90] Nour El Imane Hamda, Allel Hadjali, and Mohand Lagha. "Multisensor data fusion in IoT environments in Dempster–Shafer theory setting: An improved evidence distance-based approach". In: *Sensors* 23.11 (2023), p. 5141.
- [91] Salam Hamdan, Arafat Awaian, and Sufyan Almajali. "Compression techniques used in iot: A comparative study". In: *2019 2nd International Conference on new Trends in Computing Sciences (ICTCS)*. IEEE. 2019, pp. 1–5.
- [92] Elham Elsayed Hamza, Arabi Elsayed Keshk, and Anas Youssef. "An Experimental Network Analysis-based Approach for Detection of Jamming Attacks in Wireless Sensor Networks". In: *IJCI. International Journal of Computers and Information* 12.2 (2025), pp. 95–115.
- [93] Youkyung Han et al. "Object-based change detection of very high resolution images by fusing pixel-based change detection results using weighted Dempster–Shafer theory". In: *Remote Sensing* 12.6 (2020), p. 983.
- [94] Md. Emadul Haque et al. *A Comprehensive Survey of 5G URLLC and Challenges in the 6G Era*. 2025. arXiv: 2508.20205 [cs.NI]. URL: <https://arxiv.org/abs/2508.20205>.

- [95] Mahmudul Hasan et al. "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches". In: *Internet of Things* 7 (2019), p. 100059.
- [96] Wenhai He et al. "An Efficient Incentive Mechanism for Collaborative Anomaly Detection in Internet of Things". In: *International Conference on Parallel and Distributed Computing: Applications and Technologies*. Springer. 2024, pp. 363–371.
- [97] Jeongyoon Heo et al. "Mitigating stealthy jamming attacks in low-power and lossy wireless networks". In: *Journal of Communications and Networks* 20.2 (2018), pp. 219–230.
- [98] Martin Heusel et al. "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.
- [99] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models". In: *Advances in neural information processing systems* 33 (2020), pp. 6840–6851.
- [100] Seokhwa Hong, Kyuyeong Kim, and Si-Hyeon Lee. "A Hybrid Jamming Detection Algorithm for Wireless Communications: Simultaneous Classification of Known Attacks and Detection of Unknown Attacks". In: *IEEE Communications Letters* 27.7 (2023), pp. 1769–1773. doi: 10.1109/LCOMM.2023.3275694.
- [101] Ahmed Hussain et al. "Jamming detection in iot wireless networks: An edge-ai based approach". In: *Proceedings of the 12th International Conference on the Internet of Things*. 2022, pp. 57–64.
- [102] S. Hussain et al. "LTEInspector: A Systematic Approach for Adversarial Testing of 4G LTE". en. In: *Network and Distributed Systems Security (NDSS) Symposium 2018* (Feb. 2018). (Visited on 04/12/2024).
- [103] Syed Rafiul Hussain et al. "Privacy attacks to the 4G and 5G cellular paging protocols using side channel information". In: *Network and distributed systems security (NDSS) symposium2019* (2019). (Visited on 04/12/2024).
- [104] Samir Ifzarne, Imad Hafidi, and Nadia Idrissi. "Secure data collection for wireless sensor network". In: *Emerging Trends in ICT for Sustainable Development: the Proceedings of NICE2020 International Conference*. Springer. 2021, pp. 241–248.
- [105] Guillermo Iglesias et al. "Data augmentation techniques in time series domain: a survey and taxonomy". In: *Neural Computing and Applications* 35.14 (2023), pp. 10123–10145.
- [106] Vittorio P Illiano and Emil C Lupu. "Detecting malicious data injections in wireless sensor networks: A survey". In: *ACM Computing Surveys (CSUR)* 48.2 (2015), pp. 1–33.
- [107] Philokypros Ioulianos et al. "A signature-based intrusion detection system for the internet of things". In: *Information and Communication Technology Form* (2018).
- [108] Muhammad Irfan, Savio Sciancalepore, and Gabriele Oligeri. "Preventing Radio Fingerprinting through Low-Power Jamming". In: *Proceedings of the 20th ACM Asia Conference on Computer and Communications Security*. ASIA CCS '25. New York, NY, USA, 2025, pp. 1114–1126. doi: 10.1145/3708821.3733862.
- [109] Ali Issa, Nahi Kandil, and Nadir Hakem. "Deep Learning-based Anomaly Detection for 5G Core Mobility Management". In: *2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*. IEEE. 2023, pp. 25–29.
- [110] Nevena Jeftenić, Mitar Simić, and Zoran Stamenković. "Impact of Environmental Parameters on SNR and RSS in LoRaWAN". In: *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. IEEE. 2020, pp. 1–6.

- [111] Shashank Jere et al. "Bayesian Inference-assisted Machine Learning for Near Real-Time Jamming Detection and Classification in 5G New Radio (NR)". In: *IEEE Transactions on Wireless Communications* (2023).
- [112] Chunxiao Jiang et al. "Security in space information networks". In: *IEEE communications magazine* 53.8 (2015), pp. 82–88.
- [113] Xi Jiang et al. "NetDiffusion: Network Data Augmentation Through Protocol-Constrained Traffic Generation". In: *Proc. ACM Meas. Anal. Comput. Syst.* 8.1 (Feb. 2024). doi: 10.1145/3639037. URL: <https://doi.org/10.1145/3639037>.
- [114] John Kanwar et al. "Jamsense: Interference and jamming classification for low-power wireless networks". In: *2021 13th IFIP Wireless and Mobile Networking Conference (WMNC)*. IEEE. 2021, pp. 9–16.
- [115] Aimad Karkouch et al. "Data quality in internet of things: A state-of-the-art survey". In: *Journal of Network and Computer Applications* 73 (2016), pp. 57–81. ISSN: 1084-8045.
- [116] Parvinder Kaur, Baljit Singh Khehra, and Er Bhupinder Singh Mavi. "Data augmentation for object detection: A review". In: *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE. 2021, pp. 537–543.
- [117] Ansam Khraisat et al. "Survey of intrusion detection systems: techniques, datasets and challenges". In: *Cybersecurity* 2.1 (2019), pp. 1–22.
- [118] Gyungmin Kim and Hyuk Lim. "Reinforcement learning based beamforming jammer for unknown wireless networks". In: *IEEE Access* 8 (2020), pp. 210127–210139.
- [119] Junbong Kim et al. "GAN-Based Anomaly Detection In Imbalance Problems". In: *Computer Vision – ECCV 2020 Workshops*. Ed. by Adrien Bartoli and Andrea Fusiello. Cham: Springer International Publishing, 2020, pp. 128–145. ISBN: 978-3-030-65414-6.
- [120] Kangseok Kim. "GAN based Augmentation for Improving Anomaly Detection Accuracy in Host-based Intrusion Detection Systems". In: *International Journal of Engineering Research and Technology* 13 (Nov. 2020), p. 3987. doi: 10.37624/IJERT/13.11.2020.3987-3996.
- [121] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).
- [122] Idrees S Kocher et al. "Threat models and security issues in wireless sensor networks". In: *International Conference on Intelligent Network and Computing (ICINC)*. IACSIT. 2010, pp. VI384–389.
- [123] Katharina Kohls et al. "Lost traffic encryption: fingerprinting LTE/4G traffic on layer two". In: *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*. WiSec '19. ACM, May 2019, pp. 249–260. ISBN: 978-1-4503-6726-4. (Visited on 04/12/2024).
- [124] Don Koks and Subhash Challa. *An introduction to Bayesian and Dempster-Shafer data fusion*. DSTO Systems Sciences Laboratory Laverton Ave, ACT, Australia, 2003.
- [125] Sirisha Rani Kolli, Md Zia Ur Rahman, and Masreshaw Demelash Bayleyegn. "EECS-GT: Energy-Efficient Collaborative Sensing Model Using Game Theory for Wireless Sensor Networks". In: *IEEE Access* 12 (2024), pp. 115657–115671.
- [126] Stefan Köpsell et al. *Open-RAN Risk Analysis*. Tech. rep. Feb. 2022.
- [127] Konstantinos Kousias et al. "A Large-Scale Dataset of 4G, NB-IoT, and 5G Non-Standalone Network Measurements". In: *IEEE Communications Magazine* 62.5 (2024), pp. 44–49. doi: 10.1109/MCOM.011.2200707.

- [128] Paweł Kryszkiewicz and Marcin Hoffmann. "Open RAN for detection of a jamming attack in a 5G network". In: *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. 2023, pp. 1–2. doi: 10.1109/VTC2023-Spring57618.2023.10201067.
- [129] Paweł Kułakowski et al. "Angle-of-arrival localization based on antenna arrays for wireless sensor networks". In: *Computers & Electrical Engineering* 36.6 (2010), pp. 1181–1186.
- [130] Sean Kulinski, Saurabh Bagchi, and David I Inouye. "Feature shift detection: Localizing which features have shifted via conditional distribution tests". In: *Advances in neural information processing systems* 33 (2020), pp. 19523–19533.
- [131] Uday Kulkarni et al. "AI model compression for edge devices using optimization techniques". In: *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI, Volume 2*. Springer, 2021, pp. 227–240.
- [132] Hee-Yong Kwon, Taesic Kim, and Mun-Kyu Lee. "Advanced intrusion detection combining signature-based and behavior-based detection methods". In: *Electronics* 11.6 (2022), p. 867.
- [133] Sujitha Lakshminarayana, Amit Praseed, and P. Santhi Thilagam. "Securing the IoT Application Layer From an MQTT Protocol Perspective: Challenges and Research Prospects". In: *IEEE Communications Surveys & Tutorials* 26.4 (2024), pp. 2510–2546. doi: 10.1109/COMST.2024.3372630.
- [134] Heiner Lasi et al. "Industry 4.0". In: *Business & information systems engineering* 6.4 (2014), pp. 239–242.
- [135] JooHwa Lee and KeeHyun Park. "GAN-based imbalanced data intrusion detection system". In: *Personal and Ubiquitous Computing* 25.1 (2021), pp. 121–128.
- [136] Sheng-Wei Lei and Chih-Min Chao. "A Channel Usage Probability Awareness Anti-Jamming Channel Hopping Protocol for Cognitive Radio Networks". In: *IEEE Transactions on Cognitive Communications and Networking* 10.4 (2024), pp. 1327–1338.
- [137] Łukasz Lemieszewski, Dina Hannebauer, and Grzegorz Remiszewski. "Vulnerability of Wi-Fi wireless network to signal interference". In: *Journal of Engineering* 360 1 (Nov. 2024), pp. 58–65. doi: 10.5604/01.3001.0054.8861.
- [138] Qing Li and Wade Trappe. "Detecting spoofing and anomalous traffic in wireless networks via forge-resistant relationships". In: *IEEE Transactions on Information Forensics and Security* 2.4 (2007), pp. 793–808.
- [139] Wenjia Li, Houbing Song, and Feng Zeng. "Policy-Based Secure and Trustworthy Sensing for Internet of Things in Smart Cities". In: *IEEE Internet of Things Journal* 5.2 (2018), pp. 716–723. doi: 10.1109/JIOT.2017.2720635.
- [140] Xiaocui Li et al. "Energy-efficient anomaly detection with primary and secondary attributes in edge-cloud collaboration networks". In: *IEEE Internet of Things Journal* 8.15 (2021), pp. 12176–12188.
- [141] Yi Li et al. "Accurate anomaly detection with energy efficiency in IoT-Edge-Cloud collaborative networks". In: *IEEE Internet of Things Journal* 10.19 (2023), pp. 16959–16974.
- [142] Yuchen Li et al. "Jamming Detection and Classification in OFDM-Based UAVs via Feature- and Spectrogram-Tailored Machine Learning". In: *IEEE Access* 10 (2022), pp. 16859–16870. doi: 10.1109/ACCESS.2022.3150020.

- [143] Jianyi Lin. “Sparse models for machine learning”. In: *Engineering Mathematics and Artificial Intelligence*. CRC Press, 2023, pp. 107–146.
- [144] Fan Liu et al. “Privacy-preserving synthetic data generation for recommendation systems”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2022, pp. 1379–1389.
- [145] Hongbo Liu et al. “Topology-based multi-jammer localization in wireless networks”. In: *Security and Safety* 3 (2024), p. 2023025.
- [146] Zikun Liu et al. “Exploring practical vulnerabilities of machine learning-based wireless systems”. In: *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. 2023, pp. 1801–1817.
- [147] Yun Long, Xueyuan She, and Saibal Mukhopadhyay. “Hybridnet: integrating model-based and data-driven learning to predict evolution of dynamical systems”. In: *Conference on robot learning*. PMLR. 2018, pp. 551–560.
- [148] Howard Luu et al. “MIF: Optimizing Information Freshness in Intermittently Connected Sensor Networks”. In: *Proceedings of the 15th International Conference on Underwater Networks & Systems*. 2021, pp. 1–5.
- [149] Jianhui Ma et al. “Jamming Modulation: An Active Anti-Jamming Scheme”. In: *IEEE Transactions on Wireless Communications* 22.4 (2023), pp. 2730–2743. doi: 10.1109/TWC.2022.3213572.
- [150] Jianhui Ma et al. “Jamming modulation: An active anti-jamming scheme”. In: *IEEE Transactions on Wireless Communications* 22.4 (2022), pp. 2730–2743.
- [151] Renjie Ma et al. “Adversarial FDI attack monitoring: Toward secure defense of industrial electronics”. In: *IEEE Industrial Electronics Magazine* 18.2 (2023), pp. 48–57.
- [152] Mina Eshak Magdy et al. “Anomaly-based intrusion detection system based on feature selection and majority voting”. In: *Indonesian Journal of Electrical Engineering and Computer Science* 30.3 (2023), pp. 1699–1706.
- [153] Johan Magnusson. *Impact of Environmental Conditions on WiFi Sensing*. 2025.
- [154] Mohammed Mahyoub et al. “Security analysis of critical 5g interfaces”. In: *IEEE Communications Surveys & Tutorials* 26.4 (2024), pp. 2382–2410.
- [155] Ramona Marfievici et al. “How environmental factors impact outdoor wireless sensor networks: a case study”. In: *2013 IEEE 10th international conference on mobile ad-hoc and sensor systems*. IEEE. 2013, pp. 565–573.
- [156] Tehseen Mazhar et al. “Analysis of IoT security challenges and its solutions using artificial intelligence”. In: *Brain sciences* 13.4 (2023), p. 683.
- [157] Mehdi Mirza and Simon Osindero. “Conditional generative adversarial nets”. In: *arXiv preprint arXiv:1411.1784* (2014).
- [158] Muhammad Ahmed Mohsin et al. “Towards 6G Intelligence: The Role of Generative AI in Future Wireless Networks”. In: *arXiv preprint arXiv:2508.19495* (2025).
- [159] Paul Monferran et al. “Machine Learning Techniques for the Geolocalization of Jamming Sources in Indoor Wireless Networks”. In: *2024 International Symposium on Electromagnetic Compatibility–EMC Europe*. IEEE. 2024, pp. 534–539.
- [160] E Montcalm. *How to avoid ethical and legal issues in wireless network discovery*. 2002.

- [161] Todd Morehouse et al. "Incremental learning-based jammer classification". In: *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III*. Vol. 11746. SPIE. 2021, pp. 624–633.
- [162] Nour Moustafa and Jill Slay. "The significant features of the UNSW-NB15 and the KDD99 data sets for network intrusion detection systems". In: *2015 4th international workshop on building analysis datasets and gathering experience returns for security (BADGERS)*. IEEE. 2015, pp. 25–31.
- [163] Nour Moustafa and Jill Slay. "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)". In: *2015 military communications and information systems conference (MilCIS)*. IEEE. 2015, pp. 1–6.
- [164] Aristides Mpitzopoulos et al. "A survey on jamming attacks and countermeasures in WSNs". In: *IEEE communications surveys & tutorials* 11.4 (2009), pp. 42–56.
- [165] Hichem Mrabet et al. "A Survey of IoT Security Based on a Layered Architecture of Sensing and Data Analysis". In: *Sensors* 20.13 (2020). issn: 1424-8220. doi: 10.3390/s20133625. URL: <https://www.mdpi.com/1424-8220/20/13/3625>.
- [166] Somasundaram Muthuvel, Sivakumar Rajagopal, and Shamala K Subramaniam. "Analysis of security issues in wireless body area networks in heterogeneous networks". In: *Sensors* 22.19 (2022), p. 7588.
- [167] National Telecommunications and Information Administration. *Open RAN Security Report*. Tech. rep. May 2023.
- [168] Aaraiz Naveed et al. "Application and Ethical Aspects of Machine Learning Techniques in Networking: A Review". In: *Spectrum of engineering sciences* 2.3 (2024), pp. 455–501.
- [169] Ana Nika et al. "Towards commoditized real-time spectrum monitoring". In: *Proceedings of the 1st ACM Workshop on Hot Topics in Wireless*. 2014, pp. 25–30.
- [170] Xiangyu Niu et al. "Dynamic detection of false data injection attack in smart grid using deep learning". In: *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE. 2019, pp. 1–6.
- [171] Leena Obaid et al. "Challenges of Wireless Sensor Networks and Their Solutions". In: *International Journal of Computers and Informatics* 3 (2024), pp. 102–129.
- [172] Kamal Omari, Chaimae Taoussi, and Ayoub Oukhatar. "Comparative Analysis of Under-sampling, Oversampling, and SMOTE Techniques for Addressing Class Imbalance in Phishing Website Detection." In: *International Journal of Advanced Computer Science & Applications* 16.2 (2025).
- [173] Open5GS. *Open5GS*. URL: <https://open5gs.org/>.
- [174] Felix Ott, Lucas Heublein, and Tobias Feigl. "Research Avenues for GNSS Interference Classification Robustness: Domain Adaptation, Continual Learning & Federated Learning". In: *ECML PKDD: Industry Track* (2024).
- [175] Francesco Paissan, Alberto Ancilotto, and Elisabetta Farella. "Phinets: A scalable backbone for low-power ai at the edge". In: *ACM Transactions on Embedded Computing Systems* 21.5 (2022), pp. 1–18.
- [176] Ivan Palamà et al. "IMSI Catchers in the wild: A real world 4G/5G assessment". In: *Computer Networks* 194 (July 2021), p. 108137. issn: 1389-1286. (Visited on 09/05/2025).
- [177] Ioannis Panitsas, Iason Ofeidis, and Leandros Tassioulas. *On-Device Multimodal Federated Learning for Efficient Jamming Detection*. 2025. arXiv: 2508.09369 [cs.NI]. URL: <https://arxiv.org/abs/2508.09369>.

- [178] DJ Parish. *Decision Fusion using Dempster-Schaffer Theory*.
- [179] Seongmin Park et al. "5G security threat assessment in real networks". In: *Sensors* 21.16 (2021), p. 5524.
- [180] Thierry Parmentelat et al. *A step towards runnable papers using R2lab*. Research Report. Inria - Sophia Antipolis ; Université Côte d'Azur, June 2019. URL: <https://hal.science/hal-02167086>.
- [181] Konstantinos Pelechrinis, Marios Iliofotou, and Srikanth V Krishnamurthy. "Denial of service attacks in wireless networks: The case of jammers". In: *IEEE Communications surveys & tutorials* 13.2 (2010), pp. 245–257.
- [182] Jiaxin Peng, Siwang Zhou, and Xingting Liu. "Volatility-based measurements allocation for distributed data storage in mobile crowd sensing". In: *IEEE Systems Journal* 17.4 (2023), pp. 6665–6675.
- [183] Rong Peng and Mihail L Sichitiu. "Angle of arrival localization for wireless sensor networks". In: *2006 3rd annual IEEE communications society on sensor and ad hoc communications and networks*. Vol. 1. Ieee. 2006, pp. 374–382.
- [184] Tri Minh Triet Pham et al. "Time to Retrain? Detecting Concept Drifts in Machine Learning Systems". In: *2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE. 2025, pp. 260–271.
- [185] Chi-Dung Phung et al. "An Open Dataset for Beyond-5G Data-driven Network Automation Experiments". In: *2022 1st International Conference on 6G Networking (6GNet)*. 2022, pp. 1–4. doi: 10.1109/6GNet54646.2022.9830292.
- [186] Camelia-M Pinteá and Petrica C Pop. "Sensitive ants for denial jamming attack on wireless sensor network". In: *International Joint Conference SOCO'13-CISIS'13-ICEUTE'13: Salamanca, Spain, September 11th-13th, 2013 Proceedings*. Springer. 2014, pp. 409–418.
- [187] Roger Piqueras Jover and Vuk Marojevic. "Security and Protocol Exploit Analysis of the 5G Specifications". en. In: *IEEE Access* 7 (2019), pp. 24956–24963. ISSN: 2169-3536. (Visited on 01/04/2024).
- [188] Hossein Pirayesh, Pedram Kheirkhah Sangdeh, and Huacheng Zeng. "Securing Zig-Bee communications against constant jamming attack using neural network". In: *IEEE Internet of Things Journal* 8.6 (2020), pp. 4957–4968.
- [189] Hossein Pirayesh and Huacheng Zeng. "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey". In: *IEEE communications surveys & tutorials* 24.2 (2022), pp. 767–809.
- [190] Cong Pu. "Jamming-resilient multipath routing protocol for flying ad hoc networks". In: *IEEE Access* 6 (2018), pp. 68472–68486.
- [191] Oscar Puñal et al. "Machine learning-based jamming detection for IEEE 802.11: Design and experimental evaluation". In: *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*. 2014, pp. 1–10. doi: 10.1109/WoWMoM.2014.6918964.
- [192] Hanli Qiao, Boris Novikov, and Jan Olaf Blech. "Concept drift analysis by dynamic residual projection for effectively detecting botnet cyber-attacks in IoT scenarios". In: *IEEE Transactions on Industrial Informatics* 18.6 (2021), pp. 3692–3701.
- [193] Haneya Naeem Qureshi et al. "Toward addressing training data scarcity challenge in emerging radio access networks: A survey and framework". In: *IEEE Communications Surveys & Tutorials* 25.3 (2023), pp. 1954–1990.

- [194] Katarina Radoš, Marta Brkić, and Dinko Begušić. “Recent advances on jamming and spoofing detection in GNSS”. In: *Sensors* 24.13 (2024), p. 4210.
- [195] Meenaxi M Raikar and Meena S M. “Vulnerability assessment of MQTT protocol in Internet of Things (IoT)”. In: *2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC)*. 2021, pp. 535–540. doi: 10.1109/ICSCCC51823.2021.9478156.
- [196] Yamarthi Narasimha Rao and Kunda Suresh Babu. “An imbalanced generative adversarial network-based approach for network intrusion detection in an imbalanced dataset”. In: *Sensors* 23.1 (2023), p. 550.
- [197] Vasudev Karthik Ravindran, Sharad Shyam Ojha, and Arvind Kamboj. “A Comparative Analysis of Signature-Based and Anomaly-Based Intrusion Detection Systems”. In: *International Journal of Latest Technology in Engineering, Management & Applied Science* 14.5 (2025), pp. 209–214.
- [198] Usman Raza et al. “Practical Data Prediction for Real-World Wireless Sensor Networks”. In: *IEEE Transactions on Knowledge and Data Engineering* 27.8 (2015), pp. 2231–2244. doi: 10.1109/TKDE.2015.2411594.
- [199] Saeed Ur Rehman, Shafiq Alam, and Iman T. Ardekani. “An Overview of Radio Frequency Fingerprinting for Low-End Devices”. In: *Int. J. Mob. Comput. Multimed. Commun.* 6.3 (2014), pp. 1–21. issn: 1937-9412. doi: 10.4018/IJMCMC.2014070101.
- [200] A Antonio Reshme. “Delay and energy efficient data collection in wireless sensor networks”. In: *2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC)*. IEEE. 2019, pp. 1–6.
- [201] Daniel Romero, Tien Ngoc Ha, and Peter Gerstoft. “Spoofing attack detection in the physical layer with robustness to user movement”. In: *2024 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2024, pp. 1–6.
- [202] David Rupprecht et al. “Breaking LTE on Layer Two”. In: *2019 IEEE Symposium on Security and Privacy (SP)*. ISSN: 2375-1207. May 2019, pp. 1121–1136. (Visited on 04/11/2024).
- [203] David Rupprecht et al. “On Security Research Towards Future Mobile Network Generations”. In: *IEEE Communications Surveys & Tutorials* 20.3 (2018). Number: 3 Conference Name: IEEE Communications Surveys & Tutorials, pp. 2518–2542. issn: 1553-877X. (Visited on 04/12/2024).
- [204] Josip Šabić et al. “Practical Realization of Reactive Jamming Attack on Long-Range Wide-Area Network”. In: *Sensors* 25.8 (2025), p. 2383.
- [205] M Saedi et al. “RBS-MLP: A Deep Learning based Rogue Base Station Detection Approach for 5G Mobile Networks”. In: *IEEE Transactions on Vehicular Technology* (2024).
- [206] Shavbo Salehi et al. “Smart Jamming Attack and Mitigation on Deep Transfer Reinforcement Learning Enabled Resource Allocation for Network Slicing”. In: *IEEE Transactions on Machine Learning in Communications and Networking* (2024).
- [207] Shreya Savadatti et al. “An Extensive Classification of 5G Network Jamming Attacks”. In: *Security and Communication Networks* 2024.1 (2024), p. 2883082.

- [208] Matthias Schulz et al. "Massive reactive smartphone-based jamming using arbitrary waveforms and adaptive power control". In: *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. WiSec '17. Boston, Massachusetts: Association for Computing Machinery, 2017, pp. 111–121. ISBN: 9781450350846. DOI: 10.1145/3098243.3098253. URL: <https://doi.org/10.1145/3098243.3098253>.
- [209] Savio Sciancalepore et al. "Jamming detection in low-ber mobile indoor scenarios via deep learning". In: *IEEE Internet of Things Journal* 11.8 (2023), pp. 14682–14697.
- [210] Glenn Shafer. "Dempster-shafer theory". In: *Encyclopedia of artificial intelligence* 1 (1992), pp. 330–331.
- [211] Chowdhury Shahriar et al. "PHY-layer resiliency in OFDM communications: A tutorial". In: *IEEE Communications Surveys & Tutorials* 17.1 (2014), pp. 292–314.
- [212] Mohammad Al-Shamaileh, Patricia Anthony, and Stuart Charters. "Agent-based trust and reputation model in smart IoT environments". In: *Technologies* 12.11 (2024), p. 208.
- [213] Anuraganand Sharma, Prabhat Kumar Singh, and Rohitash Chandra. "SMOTified-GAN for class imbalanced pattern classification problems". In: *Ieee Access* 10 (2022), pp. 30655–30665.
- [214] Hong Shen et al. "Multiple source localization in wireless sensor networks based on time of arrival measurement". In: *IEEE Transactions on Signal Processing* 62.8 (2014), pp. 1938–1949.
- [215] Kexin Shi et al. "Joint System Latency and Data Freshness Optimization for Cache-enabled Mobile Crowdsensing Networks". In: *arXiv preprint arXiv:2501.14367* (2025).
- [216] Yuanming Shi et al. "Communication-efficient edge AI: Algorithms and systems". In: *IEEE communications surveys & tutorials* 22.4 (2020), pp. 2167–2191.
- [217] Yi-Sheng Shiu et al. "Physical layer security in wireless networks: A tutorial". In: *IEEE wireless Communications* 18.2 (2011), pp. 66–74.
- [218] Ankush Singla et al. "Look Before You Leap: Secure Connection Bootstrapping for 5G Networks to Defend Against Fake Base-Station". In: *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*. ASIA CCS '21. ACM, June 2021, pp. 501–515. ISBN: 978-1-4503-8287-8. (Visited on 04/11/2024).
- [219] Paweł Skokowski et al. "Jamming and jamming mitigation for selected 5G military scenarios". In: *Procedia Computer Science* 205 (2022), pp. 258–267.
- [220] Kaleb E Smith and Anthony O Smith. "Conditional GAN for timeseries generation". In: *arXiv preprint arXiv:2006.16477* (2020).
- [221] Naeimeh Soltanieh et al. "A Review of Radio Frequency Fingerprinting Techniques". In: *IEEE Journal of Radio Frequency Identification* 4.3 (2020), pp. 222–233. DOI: 10.1109/JRFID.2020.2968369.
- [222] Zineb Sordo et al. "Synthetic Scientific Image Generation with VAE, GAN, and Diffusion Model Architectures". In: *Journal of Imaging* 11.8 (2025), p. 252.
- [223] J Sormayli et al. "Real-Time jamming detection using windowing and hybrid machine learning models for pre-saturation alerts". In: *Scientific Reports* 15.1 (2025), p. 24748.
- [224] *Spectrum + Signal Analyzers - Used Keysight Equipment*. URL: <https://www.keysight.com/used/us/en/spectrum-signal-analyzers>.
- [225] srsRAN. *srsRAN*. URL: <https://github.com/srsran>.

- [226] Biljana Risteska Stojkoska and Zoran Nikolovski. "Data compression for energy efficient IoT solutions". In: *2017 25th telecommunication forum (TELFOR)*. IEEE. 2017, pp. 1–4.
- [227] May Stow. "WHEN DATA AUGMENTATION HURTS: A SYSTEMATIC EVALUATION OF SMOTE-BASED TECHNIQUES ON MEDICAL DATASETS". In: *international journal of advanced research in computer science* 16 (Aug. 2025), pp. 14–33. doi: 10.26483/ijarcs.v16i4.7313.
- [228] Luka Strizic. *Crowdsourcing GNSS Jamming Detection and Localization*. 2017.
- [229] Tomoyasu Suzaki et al. "Twine: A lightweight, versatile block cipher". In: *ECRYPT workshop on lightweight cryptography*. Vol. 2011. Springer Berlin, Heidelberg. 2011.
- [230] Bin Tang et al. "A diffusion model based on network intrusion detection method for industrial cyber-physical systems". In: *Sensors* 23.3 (2023), p. 1141.
- [231] Pietro Tedeschi, Gabriele Oliveri, and Roberto Di Pietro. "Localization of a power-modulated jammer". In: *Sensors* 22.2 (2022), p. 646.
- [232] Vishal A. Thakor, Mohammad Abdur Razzaque, and Muhammad R. A. Khandaker. "Lightweight Cryptography Algorithms for Resource-Constrained IoT Devices: A Review, Comparison and Research Opportunities". In: *IEEE Access* 9 (2021), pp. 28177–28193. doi: 10.1109/ACCESS.2021.3052867.
- [233] Okamura Toshihiko. "Lightweight cryptography applicable to various IoT devices". In: *NEC Tech. J* 12.1 (2017), pp. 67–71.
- [234] Cristian Turetta et al. "Environmental Microchanges in WiFi Sensing". In: *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE. 2024, pp. 1–2.
- [235] Amadi G Udu et al. "Emerging SMOTE and GAN-variants for Data Augmentation in Imbalance Machine Learning Tasks: A Review". In: *IEEE Access* (2025).
- [236] Bikalpa Upadhyaya, Sumei Sun, and Biplab Sikdar. "Machine Learning-based Jamming Detection in Wireless IoT Networks". In: *2019 IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS)*. 2019, pp. 1–5. doi: 10.1109/VTS-APWCS.2019.8851633.
- [237] Satish Vadlamani et al. "Jamming attacks on wireless networks: A taxonomic survey". In: *International Journal of Production Economics* 172 (2016), pp. 76–94.
- [238] Nalam Venkata Abhishek and Mohan Gurusamy. "JaDe: Low Power Jamming Detection Using Machine Learning in Vehicular Networks". In: *IEEE Wireless Communications Letters* 10.10 (2021), pp. 2210–2214. doi: 10.1109/LWC.2021.3097162.
- [239] Sanchayan Vivekananthan. "Comparative analysis of generative models: Enhancing image synthesis with vaes, gans, and stable diffusion". In: *arXiv preprint arXiv:2408.08751* (2024).
- [240] Antonia Vlahou et al. "Data sharing under the General Data Protection Regulation: time to harmonize law and research ethics?" In: *Hypertension* 77.4 (2021), pp. 1029–1035.
- [241] Jeffrey Voas et al. "Internet of Things (IoT) trust concerns". In: *NIST Tech. Rep* 1 (2018), pp. 1–50.
- [242] Omar Abdel Wahab. "Intrusion detection in the iot under data and concept drifts: Online deep learning approach". In: *IEEE Internet of Things Journal* 9.20 (2022), pp. 19706–19716.
- [243] Ning Wang et al. "Physical-layer security of 5G wireless networks for IoT: Challenges and opportunities". In: *IEEE internet of things journal* 6.5 (2019), pp. 8169–8181.

- [244] Sheng Wang and Chunliang Chu. "Geometry-covering jammer localization based on distance comprehension in wireless sensor networks". In: *arXiv preprint arXiv:1512.06468* (2015).
- [245] Tongxiang Wang et al. "Adaptive jammer localization in wireless networks". In: *Computer Networks* 141 (2018), pp. 17–30.
- [246] Mohamad Saalim Wani et al. "Security Vulnerabilities in 5G Non-Stand-Alone Networks: A Systematic Analysis and Attack Taxonomy". en. In: *Journal of Cybersecurity and Privacy* 4.1 (Mar. 2024). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, pp. 23–40. issn: 2624-800X. (Visited on 01/03/2024).
- [247] Mohammad Wazid et al. "Security in 5G-enabled internet of things communication: issues, challenges, and future research roadmap". In: *IEEE access* 9 (2020), pp. 4466–4489.
- [248] Lei Wei, Donghuai Gao, and Cheng Luo. "False data injection attacks detection with deep belief networks in smart grid". In: *2018 Chinese Automation Congress (CAC)*. IEEE. 2018, pp. 2621–2625.
- [249] Jinbo Wen et al. "Generative AI for Data Augmentation in Wireless Networks: Analysis, Applications, and Case Study". In: *arXiv preprint arXiv:2411.08341* (2024).
- [250] Axel Wirth. "The economics of cybersecurity". In: *Biomedical instrumentation & technology* 51.s6 (2017), pp. 52–59.
- [251] Yoo-Seung Won, Dirmanto Jap, and Shivam Bhasin. "Push for more: On comparison of data augmentation and SMOTE with optimised deep learning architecture for side-channel". In: *International Conference on Information Security Applications*. Springer. 2020, pp. 227–241.
- [252] Christopher Wong et al. "Internet of Things Anomaly Detection Enhancement using GAN-based Data Augmentation". In: *2024 17th International Conference on Signal Processing and Communication System (ICSPCS)*. 2024. doi: 10.1109/ICSPCS63175.2024.10815773.
- [253] Henk Wymeersch, Jaime Lien, and Moe Z Win. "Cooperative localization in wireless networks". In: *Proceedings of the IEEE* 97.2 (2009), pp. 427–450.
- [254] Bruno Missi Xavier et al. "Performance measurement dataset for open RAN with user mobility and security threats". In: *Computer Networks* 253 (2024), p. 110710.
- [255] Miao Xie et al. "Anomaly detection in wireless sensor networks: A survey". In: *Journal of Network and Computer Applications* 34.4 (2011). Advanced Topics in Cloud Computing, pp. 1302–1325. issn: 1084-8045. doi: <https://doi.org/10.1016/j.jnca.2011.03.004>. URL: <https://www.sciencedirect.com/science/article/pii/S1084804511000580>.
- [256] Jiali Xu and Valeria Loscri. "Leveraging UE-Level Collaborative Intelligence for Scalable Jamming Detection in 5G Networks". In: *2025 21st International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2025, pp. 793–797. doi: 10.1109/DCOSS-IoT65416.2025.00120. URL: <https://doi.ieeeecomputersociety.org/10.1109/DCOSS-IoT65416.2025.00120>.
- [257] Jiali Xu et al. "GANSec: Enhancing Supervised Wireless Anomaly Detection Robustness through Tailored Conditional GAN Augmentation". In: *ESORICS 2025 - 30th European Symposium on Research in Computer Security*. Toulouse, France, Sept. 2025. URL: <https://hal.science/hal-05137717>.

- [258] Jiali Xu et al. "SHIELD: Scalable and Holistic Evaluation Framework for ML-Based 5G Jamming Detection". In: *Availability, Reliability and Security*. Cham: Springer Nature Switzerland, 2025, pp. 235–256. ISBN: 978-3-032-00624-0.
- [259] Wenyuan Xu, Wade Trappe, and Yanyong Zhang. "Anti-jamming timing channels for wireless networks". In: *Proceedings of the first ACM conference on Wireless network security*. 2008, pp. 203–213.
- [260] Linjie Yan et al. "A sparse learning approach to the detection of multiple noise-like jammers". In: *IEEE Transactions on Aerospace and Electronic Systems* 56.6 (2020), pp. 4367–4383.
- [261] Yahan Yang et al. "Memory-based distribution shift detection for learning enabled cyber-physical systems with statistical guarantees". In: *ACM Transactions on Cyber-Physical Systems* 8.2 (2024), pp. 1–28.
- [262] JinYi Yoon et al. "VersatileFL: Volatility-Resilient Federated Learning in Wireless Edge Networks". In: *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE. 2023, pp. 294–302.
- [263] John Yoon. "Trustworthiness of Dynamic Moving Sensors for Secure Mobile Edge Computing". en. In: *Computers* 7.4 (Dec. 2018). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 63. ISSN: 2073-431X. (Visited on 05/04/2023).
- [264] Noe Marcelo Yungaicela-Naula, Cesar Vargas-Rosales, and Jesus Arturo Perez-Diaz. "SDN-Based Architecture for Transport and Application Layer DDoS Attack Detection by Using Machine and Deep Learning". In: *IEEE Access* 9 (2021), pp. 108495–108512. doi: 10.1109/ACCESS.2021.3101650.
- [265] Fatima Tu Zahra, Yavuz Selim Bostanci, and Mujdat Soyturk. "Real-time jamming detection in wireless IoT networks". In: *IEEE Access* 11 (2023), pp. 70425–70442.
- [266] Jing Zhang et al. "Localization for jamming attack in wireless sensor networks". In: *Intelligent Data Analysis and Applications: Proceedings of the Second Euro-China Conference on Intelligent Data Analysis and Applications, ECC 2015*. Springer. 2015, pp. 361–369.
- [267] Junwei Zhang et al. "Survey of technology in network security situation awareness". In: *Sensors* 23.5 (2023), p. 2608.
- [268] Liyang Zhang et al. "Jam sessions: Analysis and experimental evaluation of advanced jamming attacks in MIMO networks". In: *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*. 2019, pp. 61–70.
- [269] Hui Zhao et al. "Intelligent anti-jamming communication technology with electromagnetic spectrum feature cognition". In: *PloS one* 20.4 (2025), e0319953.
- [270] Tamara Zhukabayeva, Vasily Desnitsky, and Assel Abdildayeva. "Wireless Sensor Network Modeling and Analysis for Attack Detection". In: *Computer Modeling in Engineering & Sciences* 144.2 (2025), p. 2591.

Glossary

Numbers | [A](#) | [B](#) | [C](#) | [E](#) | [G](#) | [I](#) | [K](#) | [L](#) | [M](#) | [N](#) | [O](#) | [R](#) | [S](#) | [T](#) | [U](#) | [Z](#)

Numbers

3GPP The 3rd Generation Partnership Project (3GPP) is an umbrella term for a number of standards organizations which develop protocols for mobile telecommunications.. i, 12

5GC 5G Core (5GC) is a fundamental cornerstone of 5G networks establishing reliable connectivity for end users and providing access to its services. The core domain handles a range of essential functions in the mobile network, such as connectivity and mobility management, authentication and authorization, subscriber data management and policy management.. i, 132, 133

A

ADB Android Debug Bridge (adb) is a versatile command-line tool that lets you communicate with a device. The adb command facilitates a variety of device actions, such as installing and debugging apps. adb provides access to a Unix shell that you can use to run a variety of commands on a device.. i, xv, 41

AT The Hayes command set (also known as the AT command set) is a specific command language originally developed by Dale Heatherington and Dennis Hayes for the Hayes Smartmodem in 1981. The command set consists of a series of short text strings which can be combined to produce commands for operations such as dialing, hanging up, and changing the parameters of the connection. The vast majority of dial-up modems use the Hayes command set in numerous variations.. i, 46

B

bladeRF 2.0 micro The bladeRF 2.0 micro is the next-generation 2x2 MIMO, 47MHz to 6GHz frequency range, off-the-shelf USB 3.0 SDR that provides a powerful waveform development platform expected by industry professionals.. i

C

CSI Channel State Information (CSI) refers to known properties of a communication link. This information describes how a signal propagates from the transmitter to the receiver and represents the combined effect of scattering, fading, and power decay with distance. CSI provides a rich, fine-grained data source for applications like wireless sensing, localization, and security.. i

E

eMBB Enhanced Mobile Broadband (eMBB) is one of the three main use cases defined for 5G. It focuses on providing significantly higher data rates, increased capacity, and enhanced user experience for applications like high-definition video streaming, virtual reality, and augmented reality.. i

EPC Evolved Packet Core (EPC) is the core network architecture for the 4G Long-Term Evolution (LTE) standard. It provides a flat, all-IP-based architecture to handle data and voice services efficiently, serving as the predecessor to the 5GC.. i, 133

G

gNB A Next-Generation Node B (gNB) is the 5G-equivalent of a cellular base station. It is a key component of the 5G RAN and is responsible for managing radio resources and connecting UE devices to the 5GC.. i, 133

I

iPerf3 iPerf3 is a tool for active measurements of the maximum achievable bandwidth on IP networks. It supports tuning of various parameters related to timing, buffers and protocols (TCP, User Datagram Protocol (UDP), SCTP with IPv4 and IPv6). For each test it reports the bandwidth, loss, and other parameters.. i, 41, 46

K

Kurtosis In probability theory and statistics, kurtosis refers to the degree of “tailedness” in the probability distribution of a real-valued random variable. i

L

LoRa LoRa (Long Range) is a wireless communication technology designed for sending very small amounts of data over exceptionally long distances (several miles/kilometers in rural areas) while consuming minimal power, making it ideal for battery-operated Internet of Things (IoT) applications like smart agriculture, asset tracking, and city-wide sensor networks.. i, 13

M

MATLAB MATLAB is a programming and numeric computing platform used by millions of engineers and scientists to analyze data, develop algorithms, and create models.. i, 57

MQTT MQTT (Message Queuing Telemetry Transport) is a lightweight, publish-subscribe network protocol designed for low-power, high-latency, or unreliable networks, making it the perfect messaging standard for IoT devices. Its simplicity and minimal bandwidth requirements are why it’s the universal language for connecting diverse smart devices, from sensors to cloud platforms.. i, 13

N

NFV Network Functions Virtualization (NFV) is a network architecture concept that uses virtualization technologies to replace dedicated, hardware-based network appliances (e.g., routers, firewalls, load balancers) with software-based virtual machines. It is often used in conjunction with SDN to build flexible and scalable networks.. i

ns-3 ns-3 is a discrete-event, open-source network simulator primarily used for research and educational purposes. It’s designed to simulate the behavior of computer networks,

allowing users to model everything from simple LANs to complex wireless and mobile networks without needing a physical testbed.. i, 5, 28, 56, 57

NSA 5G Non-Standalone (NSA) is a deployment model for 5G networks where the 5G New Radio (NR) technology is used for the air interface, but the network management and service functions still rely on the existing 4G EPC. It is often used as an initial step in the migration to a full 5G system.. i, 133

NSSAI Network Slice Selection Assistance Information (NSSAI) is a signaling information element used in 5G networks. It helps the UE to select a specific Network Slice, which is a virtual, end-to-end network tailored to provide a certain type of service or meet the needs of a particular application.. i

O

O-RAN Open Radio Access Network (O-RAN) is an industry initiative focused on creating a more open, intelligent, virtualized, and fully interoperable RAN. It promotes disaggregation by defining open interfaces between components, aiming to increase vendor diversity and innovation.. i

R

R² R-squared, or the coefficient of determination, is a statistical measure that indicates how well a model explains the variability of the dependent variable. Its value ranges from 0 to 1, where 0 means the model explains none of the variance and 1 means it explains all the variance.. i, 51

RAN Radio Access Network (RAN) is the part of a mobile telecommunications system that implements radio access technology. Conceptually, it resides between the UE and the Core Network, providing the connection to the end user. In 5G, the RAN is primarily composed of gNB.. i, 132, 133

RFF Radio Frequency Fingerprinting (RFF) is a security technique used to identify a specific radio transmitter by analyzing the unique, minute imperfections and transient characteristics of its signal. These "fingerprints" arise from hardware manufacturing variations and can be used to authenticate devices and detect physical-layer spoofing attacks.. i

S

SA 5G Standalone (SA) is an end-to-end 5G network architecture that utilizes both the 5G RAN and the 5GC. Unlike the NSA model, it operates independently of any 4G infrastructure, enabling the full suite of 5G capabilities like network slicing and URLLC.. i

SDN Software-Defined Networking (SDN) is a network architecture approach that decouples the network control plane (which decides where traffic is sent) from the data plane (which forwards the traffic). This enables centralized management and programmability of the network, which is a foundational principle of the 5GC.. i, 132

SDR A Software-Defined Radio (SDR) is a radio communication system where components that have been traditionally implemented in hardware (e.g., mixers, filters, amplifiers, modulators/demodulators) are instead implemented by means of software on a personal computer or embedded system. This provides extreme flexibility for research, rapid prototyping, and implementing diverse communication protocols.. i

Skewness In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. i

T

TinyML Tiny Machine Learning (TinyML) is a subset of machine learning focused on the deployment of models to microcontrollers and other low-power edge devices. It brings AI to the edge of network systems, enabling real-time, low-latency, and energy-efficient inference directly on the device without relying on cloud connectivity.. i, 30, 34

U

URLLC Ultra-Reliable Low-Latency Communications (URLLC) is a key 5G service category designed to support mission-critical applications that require extremely high dependability and near-instantaneous network response. Prime use cases include industrial automation, remote surgery, and autonomous vehicle communication.. i, 133

Z

ZigBee Zigbee is a wireless communication protocol designed for low-power, low-data-rate applications, primarily used to create stable and reliable mesh networks for smart home devices like lights, sensors, and thermostats. Operating on the IEEE 802.15.4 standard, its key strength is that each device can act as a signal repeater, extending the network's range and improving its resilience.. i, 13

Contents

Abstract	xi
Acknowledgement	xiii
Acronyms	xv
Contents	xix
List of Tables	xxiii
List of Figures	xxv
Introduction	1
General Context: The 5G Era and its Security Implications	1
The Transformative Role of 5G Networks	1
A New Paradigm, A New Threat Landscape	2
The Enduring Threat of Jamming	2
The MLSysOps EU Project	3
Problem Statement and Research Gaps	3
The Scalability and Cost Barrier	4
The Robustness Gap	4
The Data Scarcity Bottleneck	5
Thesis Statement and Research Questions	5
Overview of the Proposed Methodology	6
Summary of Contributions	7
Publications and Research Outputs	8
Thesis Outline	9
1 Background and Related Work	11
1.1 The 5G and IoT Security Landscape	11
1.1.1 The 5G Architectural Paradigm and Its Inherent Vulnerabilities	12
1.1.2 The IoT Ecosystem and Resource-Constrained Security	13
1.1.3 Convergence and the Expanded Attack Surface	14
1.1.4 Protocol Layer Threats: A Case Study on Network Slicing Integrity	15
1.2 Radio Frequency Jamming as a Physical Layer Threat	16
1.2.1 Fundamentals of Jamming Attacks	17
1.2.2 A Taxonomy of Jamming Strategies	18
1.3 Paradigms for Anomalous Behaviour Detection	21

1.3.1	Signature-Based vs. Anomaly-Based Detection for Novel Threats	21
1.3.2	Limitations of Traditional and Hardware-Centric Detection Methods	22
1.3.3	Performance Evaluation of Detection Systems	24
1.4	Machine Learning Approaches for Jamming Detection	26
1.4.1	The Rationale for Machine Learning in Wireless Security	26
1.4.2	A Critical Review of Current ML-Based Detection Systems	27
1.4.3	Identified Research Gaps in Existing Literature	27
1.5	Towards Next-Generation Detection: On-Device and Collaborative Intelligence	29
1.5.1	The Imperative for On-Device Detection at the Network Edge	29
1.5.2	The Power of Collaborative Sensing in Distributed Networks	30
1.5.3	Addressing Data Scarcity: A Review of Public Datasets and Synthetic Data Generation	31
1.6	Summary and Thesis Motivation	33
2	On-Device Jamming Detection using Native System Logs	35
2.1	Motivation: A Scalable, Hardware-Agnostic Approach	35
2.1.1	Foundational Principles of Hardware-Based RF Fingerprinting	36
2.1.2	The Scalability and Economic Impasse in Massive Deployments	36
2.2	The SHIELD Framework	39
2.3	Experimental Testbed and Threat Model	40
2.3.1	Testbed Components	41
2.3.2	Threat Model	41
2.4	A Holistic Data Collection and Preprocessing Pipeline	44
2.4.1	Synchronized Data Acquisition from UE and gNB Logs	44
2.4.2	Feature Engineering from Native Logs	47
2.4.3	Handling Asynchronous and Missing Data via Interpolation	48
2.4.4	Temporal Feature Extraction via Sliding-Window Aggregation	51
2.5	A Rigorous Methodology for Evaluating Model Robustness	53
2.5.1	Analysis of UE Feature Distributions	54
2.5.2	Analysis of gNB Feature Distributions	55
2.6	Performance Evaluation	56
2.6.1	Benchmarking Existing Methods	56
2.6.2	Performance of the SHIELD Pipeline	57
2.7	Chapter Summary	59
3	Enhancing Detector Robustness with Generative Data Augmentation	61
3.1	The Challenge of Data Scarcity and Model Generalization	61
3.1.1	The Data Acquisition Bottleneck	62
3.1.2	Distribution Shift and the Failure of Generalization	65
3.1.3	The Inadequacy of Traditional Augmentation Techniques	66
3.1.4	The Case for Generative Models: A New Paradigm for Data Synthesis	69
3.2	The GANSec Framework for Wireless Time-Series Augmentation	73
3.2.1	Conditional GAN Fundamentals	73
3.2.2	Generator and Discriminator Design	74
3.2.3	Fidelity-focused vs. Task-oriented Generation	75
3.2.4	Tailoring GANSec for Wireless Data	76
3.3	Evaluation of the Synthetic-Only Training Strategy	78
3.3.1	Experimental Design and Baselines	78
3.3.2	Training Settings for Diversity-Fidelity Balance	80

3.3.3 Quantitative Analysis of Synthetic Data Quality	81
3.3.4 Cross-Scenario Evaluation of Anomaly Detector Robustness	83
3.4 Chapter Summary	85
4 A Framework for Collaborative Intelligence	87
4.1 Fundamental Limitations of On-Device Threat Detection	87
4.1.1 Inability to Assess Geographic Scope and Scale	88
4.1.2 The Inherent Incapability of Single-Point Source Localization	88
4.1.3 Resolving Ambiguity: Differentiating Malicious Attacks from Benign Faults	89
4.2 A Lightweight Collaborative Intelligence Architecture	90
4.2.1 Core Architectural Principles	90
4.2.2 System Components and Workflow	91
4.3 Fusion and Localization from Distributed Binary Data	93
4.3.1 Aggregating Unreliable Inputs with Evidence Theory	93
4.3.2 Jammer Localization using Boundary-Based Techniques	94
4.3.3 Deriving Fine-Grained Network Insights	95
4.4 In-Depth Analysis of Operational Challenges and Mitigations	97
4.4.1 Communication Overhead	97
4.4.2 Data Reliability and FDI Resilience	97
4.4.3 Impact of UE Density and Distribution	99
4.4.4 System Dynamics and UE Mobility	99
4.4.5 Handling Multiple Jamming Sources	100
4.5 Chapter Summary	100
5 Discussion and Conclusion	103
5.1 Synthesis of Contributions	103
5.2 Interpretation of Key Findings	104
5.2.1 The Efficacy of On-Device, Log-Based Detection	104
5.2.2 The Power of the Synthetic-Only Training Strategy	105
5.2.3 The Viability of Lightweight Collaborative Intelligence	106
5.3 Implications of the Research	106
5.3.1 Practical Implications	106
5.3.2 Methodological Implications	108
5.3.3 Theoretical Implications	108
5.4 Limitations of the Presented Work	109
5.4.1 Scope of Experimental Validation	109
5.4.2 Assumptions in the Generative and Collaborative Models	109
5.4.3 Lack of Quantitative Evaluation for the Collaborative Framework	110
5.5 Future Research Directions	110
5.5.1 Expanding Experimental Validation	110
5.5.2 Advancing the ML Methodologies	111
5.5.3 Hardening the Collaborative Framework	111
5.5.4 Moving from Detection to Autonomous Mitigation	112
5.6 Concluding Remarks	112
Bibliography	113
Glossary	131
Contents	135

CHARACTERISATION OF ANOMALOUS BEHAVIOUR FOR SECURITY IN DEEP-EDGE WIRELESS SYSTEMS
Applied Machine Learning From On-Device Jamming Detection to Collaborative Intelligence

Abstract

The proliferation of Fifth-generation (5G) networks is critical for modern society, yet their reliance on the open radio frequency spectrum makes them persistently vulnerable to jamming attacks. Existing detection methods often lack the scalability, cost-effectiveness, and robustness required for widespread deployment, and their performance is further hindered by the scarcity of anomalous data needed to train reliable machine learning models. This thesis addresses these challenges by proposing and validating a comprehensive, multi-stage framework for characterising and detecting anomalous behaviour, progressing from robust on-device sensing to scalable, collaborative intelligence. First, this work establishes the feasibility of a hardware-agnostic approach, demonstrating that robust jamming detection can be achieved using only native system logs from commercial User Equipment (UE) and base stations. A holistic data processing and evaluation pipeline (SHIELD) is introduced, proving its effectiveness against sophisticated, power-modulated jammers where conventional methods fail. Second, to overcome the critical limitations of data scarcity and poor model generalization, a tailored conditional Generative Adversarial Network (GANSec) is developed. A novel synthetic-only training strategy is shown to significantly enhance detector robustness, enabling models to maintain high performance under unseen network conditions. Finally, the solution is scaled to a system-wide level through a lightweight collaborative intelligence framework. This architecture leverages minimalist binary reporting from a multitude of UEs, which are aggregated to achieve network-wide situational awareness, including jammed area mapping and threat localization, with minimal communication overhead. The results of this research demonstrate a practical and effective end-to-end solution that advances the state of the art in wireless security. Serving as a foundational component of the EU-funded MLSysOps project, this work provides the core methodologies for a wireless network anomaly detection module aimed at ensuring resilience and trust in heterogeneous networks. By uniting on-device data collection with advanced machine learning for both data augmentation and collaborative sensing, this thesis presents a scalable and resilient paradigm for defending next-generation networks against physical layer threats.

Keywords: 5g security, machine learning, jamming attacks, anomaly detection, data augmentation, collaborative intelligence

Centre Inria de l'Université de Lille

Parc scientifique de la Haute-Borne – 40, avenue Halley - Bât A - Park Plaza – 59650 Villeneuve d'Ascq - France

Résumé

La prolifération des réseaux de cinquième génération (5G) est essentielle pour la société moderne, mais leur dépendance au spectre radiofréquence ouvert les rend vulnérables aux attaques par brouillage. Les méthodes de détection existantes manquent souvent de scalabilité, de rentabilité et de robustesse pour un déploiement à grande échelle, et leurs performances sont limitées par la rareté des données anormales nécessaires à l'entraînement de modèles d'apprentissage automatique fiables. Cette thèse aborde ces défis en proposant et en validant un framework complet et multi-étapes pour la caractérisation et la détection des comportements anormaux, progressant d'une détection robuste sur appareil à une intelligence collaborative et scalable. Premièrement, ce travail établit la faisabilité d'une approche agnostique du matériel, démontrant qu'une détection de brouillage robuste peut être réalisée en utilisant uniquement les journaux systèmes natifs des équipements utilisateurs (UE) et des stations de base commerciales. Un pipeline holistique de traitement et d'évaluation des données (SHIELD) est introduit, prouvant son efficacité contre des brouilleurs sophistiqués à modulation de puissance, là où les méthodes conventionnelles échouent. Deuxièmement, pour surmonter les limitations critiques de la rareté des données et de la faible généralisation des modèles, un réseau antagoniste génératif conditionnel (GANSec) est développé. Une nouvelle stratégie d'entraînement basée uniquement sur des données synthétiques améliore significativement la robustesse des détecteurs, permettant aux modèles de maintenir de hautes performances dans des conditions réseau inédites. Enfin, la solution est étendue à l'échelle du système grâce à un framework d'intelligence collaborative léger. Cette architecture s'appuie sur des rapports binaires minimalistes provenant d'une multitude d'UE, qui sont agrégés pour obtenir une connaissance situationnelle à l'échelle du réseau, incluant la cartographie de la zone brouillée et la localisation de la menace, avec une surcharge de communication minimale. Les résultats de cette recherche démontrent une solution de bout en bout, pratique et efficace, qui fait progresser l'état de l'art en matière de sécurité sans fil. Servant de composant fondamental au projet MLSysOps financé par l'UE, ce travail fournit les méthodologies de base pour un module de détection d'anomalies de réseau sans fil visant à garantir la résilience et la confiance dans les réseaux hétérogènes. En unifiant la collecte de données sur appareil avec des techniques d'apprentissage automatique avancées pour l'augmentation de données et la détection collaborative, cette thèse présente un paradigme scalable et résilient pour la défense des réseaux de nouvelle génération contre les menaces de la couche physique.

Mots clés : sécurité 5g, apprentissage automatique, attaques de brouillage, détection d'anomalies, augmentation de données, intelligence collaborative
